THREE ESSAYS ON EMPIRICAL INDUSTRIAL ORGANIZATION

BY

ZENING LI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor George Deltas, Chair
Professor Dan Bernhardt
Associate Professor Seung-Hyun Hong
Professor Dan McMillen

Abstract

This dissertation contains three chapters on topics in the field of empirical industrial organization. The first two chapters focus on lender and borrower strategies in the U.S. mortgage market while the third chapter addresses the asymmetric price adjustment phenomenon in the U.S. gasoline market.

The first chapter shows how consumer search confers positive externalities to other consumers in the same market. These externalities can be either direct, by sharing information from prior searches and thus improving the effectiveness of the search process, or indirect, by changing the equilibrium strategies of firms. Either type of externality allows consumers to reduce their own costly search activity: a consumer in a market populated with low search cost consumers searches less than an otherwise identical consumer in another market populated with high search cost consumers, while obtaining lower prices. We present evidence in support of the presence of both direct and indirect externalities in the U.S. mortgage issuance industry, though the evidence is stronger for the former than for the latter. Given that in the mortgage industry lenders tend to charge the same mortgage rates within all markets in a state, indirect externalities operate through the composition of active firms in a market. This margin is typically ignored in the theoretical literature, as often is the possibility of direct information externalities.

The second chapter investigates the systematic differences in pricing between mortgage lenders operating in many states and those operating in one or a handful of states. We provide evidence that multi-state lenders' price for a given mortgage product is influenced by the price they charge for that same product in their other markets. Moreover, the pricing of a product in a state reflects the importance of that product in all the markets a lender is active in rather than its importance in that particular state. These effects are more pronounced for non-bank lenders and for products offered to high-risk borrowers. Given the large variability in prices, most of which invisible to

borrowers across state-lines, consumer aversion to geographic price discrimination is unlikely to be a factor for "not pricing to the market". Because a lender's cost-of-funds is unlikely to vary differentially by product from that of other lenders, cost-based explanations are also unlikely. This leaves organizational and informational factors as the most likely sources of pricing differences between multi-state and local lenders.

The final chapter examines how asymmetric price adjustment speeds in the U.S. gasoline market vary across time. The 20-year sample consisting of weekly New York Harbor gasoline spot prices and U.S. retail gasoline prices from 1993 to 2013 is first studied in its entirety and then divided according to breakpoints detected by structural break tests. I use a generalized asymmetric error correction model and a time-varying coefficient model to test whether retail gasoline prices respond more quickly to increases than to decreases in spot market gasoline prices and calculate the cumulative response function along with the associated consumer costs. Estimation results not only confirm the presence of asymmetric adjustments, but also suggest the degree of asymmetry increasing across time.

To those close to my heart

Acknowledgments

First and foremost, I own a debt of gratitude to my adviser George Deltas, who shared with me his wisdom in the discipline of economics and in life. It was a privilege to work with him, and this thesis could not have been completed without his guidance and support. I am also truly thankful for the comments and encouragement I received from my committee members: Dan Bernhardt, Seung-Hyun Hong, and Dan McMillen. In addition, I am grateful to the faculty, staff, and students in the economics department for being there for me from the beginning of this journey to the very end.

I was fortunate to be accompanied with amazing friends during my Ph.D. years. They are the ones I could turn to in happiness or despair. Xiao Wang, my close friend since 12, joined me here for her Ph.D. in mathematics, and took care of me in every possible way a friend could. Rui Fan, my rock in the same cohort, offered me a different perspective on things, oftentimes more insightful and unfailingly more optimistic. Daria Bottan, with whom I've shared many memorable and profound conversations, showed me how a beautiful human being could be strong yet vulnerable.

I also wish to thank my parents, my father, for constantly pushing me out of my comfort zone, and my mother, for her unconditional love. I dedicate every bit of my success to them, knowing how pleased and proud they must be.

Finally, a special thanks to my husband, Shuo Zhang, who has contributed to the first two chapters of this thesis by generously providing his knowledge of the mortgage industry. Having him by my side, always believing in me, always cheering me on, and in the darkest of times, loving me more than I love myself, is the greatest treasure I could ever possess.

Table of Contents

Chapter 1

Free Riding on the Search of Others: Information Externalities in the U.S. Mortgage Industry

1.1 Introduction

Product prices are often not freely observable. As noted in a voluminous literature, starting with Stigler (1961), consumers may be aware only of the price distribution and must engage in costly information acquisition to obtain specific price quotes. This information acquisition often takes the form of search, where a consumer balances the cost of obtaining an additional quote, versus the benefit of discovering a lower price.[1] Consumers vary in search costs. Therefore, different consumers facing the same price distribution differ in their search activity. Because markets consist of different proportions of various consumer groups, the search activity in some markets may exceed that of other markets even if the price distribution in both markets were the same. The former set of markets could be referred to as having a high baseline search propensity, while the latter could be referred to as having a low baseline search propensity. This baseline search propensity features prominently in the analysis we undertake in this chapter.

Suppliers also often vary in their costs, and thus differ in their optimal prices under a given set of demand conditions. For each supplier, the search activity of consumers and the posted prices of competing suppliers yield a demand curve. The supplier sets a price that is a function of its cost. In the market equilibrium, the price and search distributions are such that (a) no consumer wants to change his or her search activity, (b) no supplier in the market wants to change its price,

This chapter is based on joint work with George Deltas. We would like to thank Alexei Alexandrov, Gene Amromin, Dan Bernhardt, Thomas Conkling, Seung-Hyun Hong, Sergei Koulayev, Dan McMillen, Leonard Nakamura, and seminar participants at the International Industrial Organization Conference, the Society for Economic Measurement Conference and at the University of Illinois for useful comments.

[1] Though our language may be suggestive of sequential search, we do not presume a specific search protocol. As Manning and Morgan (1982) point out, a fixed sample size (fixed n) search may often be superior, and an optimal search may contain features of both (see Morgan and Manning (1985); Wilde and Schwartz (1979) also provides some relevant discussion).

(c) no supplier who is in the market wants to exit (or a supplier who is not in the market wants to enter).[2] Most theoretical work in search takes the set of firms as given, and thus ignores the last element of the market equilibrium. However, this element is often of empirical relevance, and we will return to it shortly.

When deciding on his or her search activity, a consumer only weighs the private benefits against the private costs. But search activity also generates externalities to other consumers, as has been recognized since the early contributions of Salop and Stiglitz (1977) and Varian (1980). One part of these externalities is direct. For example, consumers may share their experience with their social network, informing their friends about low price vendors to shop at and high price vendors that are best avoided. For multi-product vendors, this information may be valuable beyond the specific product that a consumer has bought: information that a consumer has found a good price for one product in a vendor may suggest to that consumer's circle of friends that the vendor is a good prospect for purchasing other products. For example, if a borrower has secured a low interest rate car loan from a bank, some of his friends or relatives who are cognizant of this may consider that bank when searching for a mortgage. These direct externalities result in search that utilizes (imperfect) vendor-specific price information and that is partially directed rather than purely random. It can thus be more efficient, as low price vendors may be "oversampled" relative to high price vendors.[3]

Another part of the search externalities may be indirect and operate through the market equilibrium. The prevalence of low search cost (high search intensity) consumers makes a market more competitive and prices generally lower. In fact, in the benchmark sequential model of Stahl (1989) the entire price distribution shifts to the left as the proportion of informed consumers increases; similar results are present in other models. These price effects can arise either from lower equilibrium margins of the firms active in the market, or by a change in the composition of firms towards lower cost vendors, or a combination of both. In turn, these effects can reduce the incentives of consumers to undertake search activity. Indeed, as Salop and Stiglitz (1977) have pointed

[2]Non-degenerate equilibrium price distributions do not require firm and/or consumer heterogeneity, which underlies the discussion here. Burdett and Judd (1993) derive conditions under which price dispersion arises even with homogeneous consumers and firms.

[3]This is a topic that has been studied by the marketing literature, with Brown and Reingen (1987) among the early formal contributions. The idea of product information being exchanged "over the clothesline" and "across backyard fences" dates to Whyte (1954).

out, high search cost consumers confer a negative externality to low search cost consumers, by making it possible for high priced firms to remain in the market: low search cost consumers would now have to search longer until they find a low price. Therefore, a consumer in a market with a high baseline search activity may search less than an otherwise identical consumer in a market with a low baseline search.

We ask whether there is any evidence for direct or indirect search externalities in the U.S. mortgage loan industry and attempt to measure their importance. We investigate to what extent these externalities affect the search activity, the price distribution, and the set of firms active in each local mortgage market. We combine data on both search intensity and transaction prices by county for year 2014; older data are used for the construction of instruments as appropriate. We find that both types of externalities are present and non-negligible. Direct externalities appear to be stronger than the indirect ones, at least for the mortgage market. Since financial institutions tend to have uniform pricing within a state for regulatory/litigation concerns, indirect externalities manifest themselves only through the differential participation of high versus low priced lenders across geographical submarkets.

Because of the nature of the questions we ask and the type of data in our disposal, we adopt a reduced form approach. We identify the search propensity associated with each borrower, property, and loan characteristic from the within-market differences in search rates. This part of the analysis utilizes data collected on the basis of the Home Mortgage Disclosure Act (HMDA) and contains a smaller set of variables than those available in the credit application. Importantly, however, it also contains data on some variables that may capture search activity differentials that cannot directly affect mortgage rates, such as race, ethnic background, and gender. Some of the borrower characteristics have the expected association with search rates: all things equal, borrowers that qualify for VA-guaranteed loans search less than Conventional or FHA-insured loans, perhaps because VA-guaranteed loans can only be issued by qualified lenders, and the terms are usually much better to begin with; low income individuals search somewhat more (possibly because of lower opportunity cost of time relative to interest savings); those who refinance search more (there is a less time pressure to secure a loan, and this is also a selected group of borrowers). There are no prior expectations for other characteristics, but we do find differential propensities based on borrower ethnic background and gender. Finally, the loan amount has a somewhat unexpected association with search intensity: those who apply for the smallest loans search the most. A positive association might be expected, since larger loans imply larger gains from identifying a lower rate.[4] But a likely explanation is that the loan size itself may be correlated with borrower characteristics that are not included in the HMDA database, e.g., the credit score, LTV, income, which may affect the level and dispersion of rates that these borrowers are quoted.

We use the composition of borrowers in each market and the coefficients from the regression described above to construct an index of baseline search activity (henceforth *BSI*). We then regress the market search rates, adjusted for the search propensity of its constituent borrowers, on the baseline search activity index (*BSI*) and other market characteristics and proxies that may be relevant for search activity, e.g., the concentration of mortgage lending industry, median rents, etc. In the absence of any search externalities, the coefficient of *BSI* would be zero, unless it were correlated with unobserved factors that are relevant for search activity. However, the coefficient on *BSI* is consistently negative and significant. In other words, an individual in a market populated by borrowers with high search propensity is searching less himself/herself. The estimated parameter varies by specifications, but the range of estimates suggests a nearly complete crowd out, a result to which we return to below.

We next relate the baseline search activity to the price distribution faced by borrowers in each market. To do so, we utilize a database that includes the interest rates and most other relevant mortgage application characteristics from all loans securitized by Fannie Mae, Freddie Mac, and Ginnie Mae. We adjust the mortgage rate for all application characteristics to construct a residual mortgage "price". Because an institution's mortgage pricing generally does not vary within a state, the only channel for across-market differences in the offered prices comes from differences in the seller composition.[5] We show that lenders with high prices are more likely to operate in low *BSI* markets, relative to lenders with low prices. Recall that prices cannot be tailored to lo-

[4]This reasoning is reminiscent of that in Sorensen (2000), who finds that price dispersion is lower for "maintenance" drugs, presumably because consumers have greater incentives to search for low prices for drugs they will be purchasing for many years.

[5]This geographical price uniformity, at least within individual states, is a feature of the U.S. market and derives from the prohibition of "red-lining". This prohibition is often based on the premise of non-discrimination, but can also be justified on the basis of correcting privately optimal but socially suboptimal lender behavior (see Lang and Nakamura, 1993). In other countries, mortgage rates may be negotiated at the individual level (see Allen, Clark, and Houde, 2016). In areas where there is no heterogeneity in risk or information for mortgage issuance, multi-branch banks may prefer to commit to uniform pricing (see Calem and Nakamura, 1998). In this case, fear of litigation serves as a commitment device that increases equilibrium profits of multi-branch banks.

cal market conditions; the lender can only choose whether or not to operate there. As a result, markets with high values of *BSI* have marginally lower offered mortgage prices, after controlling for other relevant market characteristics. More importantly, we next compute separate distributions for mortgage price quotes (based on rates offered to mortgage applicants) and for mortgage price transactions (based on rates accepted by borrowers, i.e., based on rates of originated loans). The distribution of mortgage price quotes is noticeably more responsive to *BSI* than the average lender price in a market, and the distribution of accepted mortgage prices is even more responsive. These findings suggest that search in areas with high *BSI* is more "efficient", especially given the essentially full crowd-out effect reported in the preceding paragraph. Prospective borrowers in high *BSI* markets seem to obtain quotes from lower priced institutions; if they get multiple quotes, perhaps one from the lender they hold a deposit account (or the one the realtor recommends), any additional quote seems to be coming from relatively lower priced lenders, possibly based on information from their social network.[6]

These findings has a number of implications. The first implication is that search externalities lead to a sub-optimally low level of consumer search in the mortgage market. Search is a "public good" that aids price discovery, increases market competitiveness, and pushes inefficient producers out of the market. These effects are clearly present in the mortgage industry. Thus, policies that reduce the search cost for consumers of mortgage services would be welfare enhancing if these policies have relatively modest costs. Second, high search groups provide an implicit transfer to co-located low search groups, i.e., as noted above search externalities have a redistributive element. Lastly, the presence of direct search externalities might pose a problem for simple structural search models. A observed low transaction price is perhaps due to a "tip" from friends/family and not the outcome of extensive search. Transaction prices are lower than implied by the number of searches, unless provision is made in the model to allow for some searches to be partially informed.

This chapter is related to the empirical literature on search, most of which uses structural empirical methodologies to estimate the market fundamentals and perform counterfactual simulations. One part of this literature that is relevant to our study focuses on the mortgage industry

[6] The lender from which a particular consumer gets a quote deterministically (i.e., with probability one) is "prominent" in the terminology of Armstrong, Vickers and Zhou (2009). However, in the mortgage market, the identity of this prominent lender differs from consumer to consumer, even within a specific county.

and investigates issues of price dispersion and search frictions. Gurun, Matvos, and Seru (2013) find large differences (2.8 percentage points between the 5th and 95th percentile lenders on average) in reset rates for adjustable-rate mortgage (ARM) loans charged by lenders within geographic regions, after conditioning on borrower and loan characteristics and the initial interest rate. Alexandrov and Koulayev (2017) provide some direct evidence on the extent of mortgage price dispersion, obtained through the use of a unique proprietary data of mortgage rate sheets collected by the CFPB which covers the most important lenders. Even the most competitive segments of the market exhibit a dispersion of 0.5%, which is similar to the price dispersion we obtain in our study using a different dataset and methodology.[7] They then proceed to calculate the consumer gains if the fraction of consumers who search increased by a substantial (but plausible) amount. Allen, Clark, and Houde (2016) suggest that search frictions in the Canadian mortgage industry reduce consumer surplus by almost $12 per month per consumer and that 28% of this reduction can be associated with discrimination, 22% with inefficient matching, and the remainder with search cost. In a similar vein, Woodward and Hall (2012) point out that confused consumers overpay their brokers' services at least $1,000 by shopping from too few brokers. Perhaps due to its financial complexity, searching for a suitable mortgage does seem to be a daunting task for most consumers, a finding that is implicit in our work.

A second relevant component of the empirical search literature estimates the direct effects from consumer search (i.e., the effects of increased search on transaction prices for a given price distribution) and contrasts them with the strategic effects of search (i.e., the effects that increased search has on the price distribution). This latter effect is what we refer to as indirect externalities. Recent work by Brown (2018) shows that for relatively small search levels, such as those for medical imaging procedures in New Hampshire, the direct gains from search dominate the strategic ones, but for high levels of search the reverse is true. Another recent contribution, by Salz (2015), finds evidence that in the New York trade waste market, those who use intermediaries (brokers) to identify low price providers confer an externality to those who do not use intermediaries. This finding explicitly links the effects of increased search on the price distribution to a redistribution of surplus from the firms to the consumers who choose not to search.

[7]Price dispersion in homogeneous financial products is not limited to mortgages, as Hortacsu and Syverson (2003) point out in their study of mutual funds that track the S&P index.

1.2 Data

The data we use come from two main sources and are also complemented with information from other data series. The first source is the annual loan application register data provided by the Home Mortgage Disclosure Act (HMDA) of 1975, which requires many depository and non-depository lenders to collect and publicly disclose information about housing-related loans and applications. Whether a lender satisfies HMDA's coverage criteria depends on its size, the extent of its business in an MSA, and whether it is in the business of residential mortgage lending.[8] In 2014, 7,062 institutions reported data on nearly 10 million home mortgage applications, which is the vast majority of all applications made during this time.

Lenders report to HMDA the action taken following a loan application. Table 1.1 Panel A lists the number of observations corresponding to each of these actions for first-lien, one-to-four family home mortgage applications.[9] Of all the applications in 2014, nearly 61% resulted in an origination. The second largest action category (17.71%) is denial of the application by the lender. Denied applications are not considered as searches for the purpose of our study because they do not yield an option to borrow funds. Moreover, applicants who get turned down by a lender during their first mortgage application typically do not apply to another institution.[10] For the same reason, incomplete applications are also not considered valid searches. Applications approved but not accepted and applications withdrawn by the applicant sum up to more than 15% of the total applications.[11] These two actions are categorized as searches that borrowers made, and after comparison with their other options, chose to turn down.[12]

HMDA also contains information on borrower characteristics, such as the applicant's race, gender, ethnicity, income, and also information on loan characteristics, such as the loan amount

[8]Specific regulatory disclosure requirements can be found at: https://www.ffiec.gov/hmda/pdf/2013guide.pdf

[9]The current version of 2014 HMDA data that is downloadable from the CFPB website is slightly different from what we downloaded in 2016 in terms of the transaction numbers. This is because the 2014 data were updated to include approximately 174,000 transactions from Green Tree Servicing, LLC, which were not incorporated into the 2014 data until after the dataset was finalized.

[10]See Mondragon (2015). The reason for this behavior may be that the applicant infers from the denial that they are not creditworthy and will also be turned down by other institutions.

[11]The top 2 reasons for withdrawing an application is that the borrower obtained a better quote or more timely funding from other sources (http://linear-title.com/top-reasons-a-borrower-might-withdraw-their-mortgage-application/). Changes in the borrower's circumstances are another common driver for application withdrawals.

[12]HMDA also has limited data on pre-approval requests. This data is not used because doing so requires non-trivial adjustments to account for the incompleteness of the data.

and purpose, owner occupancy, loan type, property type, and lender ID. Nearly all of the applications are available at the census tract level, where the location refers to that of the property, but we use county level data in our analysis. The main reason is that the census tract is too narrow a geographical area for defining retail mortgage markets. On the other hand, the county is a reasonable geographical partition for market definition. Amel, Kennickell, and Moore (2008) find that the median household lived within four miles of its primary financial institution and 25% of households obtained mortgages from their primary financial institution. In addition, more than 50% of households obtained mortgages from an institution less than 25 miles away. This is a high percentage, in light of the fact that a substantial fraction of mortgage sellers are online firms operating nationally. This figure indicates that the majority of borrowers search at least one local mortgage lender if not more. They might still use other resources to obtain additional information or price quotes, but it seems like they tend to originate their loan with a local lender. Borrower location will sometimes not correspond to property location, which is the geographic information in our data, because many home purchases take place when people move across counties; this fact further explains why a portion of the mortgages are obtained from lenders that are not in close proximity to the borrower.

Note that given the nature of the geographic identifier in HMDA, we define markets based on the location of properties, not the location of borrowers. Of course, since the properties in our sample are for most part owner occupied properties, following the purchase owners and properties will typically be in the same location; but this might not be the case when the mortgage application is filed. However, a person who is moving across counties and who obtains financing from an institution in the destination county will still be subject to the indirect externalities of that county to the exact same extent as a resident and will also be subject to the direct externalities to the extent that he or she has personal contacts in the destination county.[13]

We show our data selection process in Table 1.1 Panel B. Starting with close to 9 million mortgage applications for first-lien, 1-4 family homes, we first drop applications that are not categorized as searches, i.e. applications that are denied by the financial institution, closed due to incompleteness, or which were pre-approval requests. This leaves us with approximately 6.8 million

[13]If that person were to obtain financing from his county of origin, then the externalities would be those of that county, but our procedure will (incorrectly) assign to him or her the externalities of the destination county.

searches in 2014. Since our analysis is at the county level, we then drop applications with county code missing. We also drop applications for which race, gender, ethnicity, or owner-occupancy questions are "not applicable", which usually indicates that the applicant is not a natural person (for example, a corporation). Since the loan terms, approval criteria, loan characteristics, etc. for commercial mortgage loans are very different from residential mortgage loans, we consider them as mortgage products in two separate markets. Moreover, it is unclear how, or if, search behavior and externalities in the individual borrower residential mortgage loan market would relate to applications in the business sector, making their inclusion conceptually problematic. After the above selection process, we have 6.7 million applications remaining in the sample.

Of these applications, approximately a fifth did not result in originations. We consider those to be a measure of borrower search activity. It may appear striking that the number of searches per "purchase" is only 1.25, so some discussion is warranted. One reason for this low figure is that our definition of a search is rather stringent. For example, browsing the web to identify an institution's headline rates or to obtain a quote based on partial borrower information is not considered a search, largely because these rates may be weakly related to the actual rates a borrower would obtain if he/she completed an application. However, we recognize that scouring the web for quotes has some value, as do pre-approval requests. Given this, the most appropriate way to interpret our measure of search is that it is a proxy for overall search activity, since the number of these other "softer" quotes that borrowers obtain are likely strongly correlated with the mortgage applications they file to obtain hard quotes. Importantly, in our interpretation of the results, we will also consider our measure of search to be a proxy for the extent of information acquisition from the borrowers' social network, which would provide direct recommendations and information about the lender's reputation.[14] The second reason why the average search figure is low is because prospective borrowers do not, in fact, engage in much search activity. According to the National Survey of Mortgage Borrowers (NSMB), almost half of consumers who take out a mortgage for home purchase fail to shop around: they seriously consider only one single lender

[14]In the 2015 study by the CFPB on *Consumers' Mortgage Shopping Experience*, "reputation of the lender/broker" and "recommendation from a friend/relative/co-worker" were among the top five considerations when shopping from a particular lender. This report is available at http://www.consumerfinance.gov/reports/consumers-mortgage-shopping-experience. According to Lacko and Pappalardo (2007), in a survey conducted by the Federal Trade Commission, almost all respondents collected mortgage information from newspapers or the Internet and many of the respondents who did not contact multiple lenders relied on recommendations from friends.

before choosing where to apply. Moreover, for most borrowers, their mortgage shopping experience stops with their first application, as corroborated in the HMDA data.[15] These low search rates are despite the large dispersion in mortgage rates across financial institutions, even after accounting for loan size and mortgage type, as recently documented by the Consumer Financial Protection Bureau (CFPB)'s analysis of mortgage rate quotes.[16]

The HMDA dataset does not link all applications filed by the same individual, i.e., there are no borrower identifiers. We thus group observations based on combinations of observable characteristics (described in detail below). This grouping rests on the following two assumptions. First, we assume that borrowers who share the same observable characteristics across locations have similar search propensities, even though they will have different equilibrium search rates. That is, if these borrowers were placed in identical environments, their search activity would be similar; actual search activity will differ only because of market characteristics, e.g., the distribution of mortgage rates. Search "primitives" such as search costs, prior information about the mortgage market, and gains from incremental search depend similarly on observable borrower and property characteristics (and associated proxies) across markets.

Second, we assume that the bulk of applicants primarily search (for both properties and mortgages) within a market, report the same characteristics in all applications, consider properties and loans of a given type, and originate the loan within the calendar year if they were approved. Therefore, the vast majority of applications within a market that share the same borrower, loan, and property characteristics will be filed by individuals with the same search propensity. Their search propensity will be the same, given the assumption that it is a function of borrower, property, and loan characteristics, which results in the same equilibrium search intensity since these individuals are active in the same mortgage market.[17] The expected search activity in a particular

[15]See *Consumers' Mortgage Shopping Experience* (CFPB, 2015), chapters 2 and 3, for relevant figures.

[16]The CFPB notes that "shopping is important not only to help borrowers understand the different product features available, such as adjustable-rate versus fixed-rate, but also the price at which those products are offered." Recognizing the potential benefits of effective shopping, the CFPB is improving mortgage disclosures under the Truth in Lending Act and the Real Estate Settlement Procedures Act. In October 2015, the "Know Before You Owe" mortgage disclosure rule replaces four disclosure forms with two new ones, the Loan Estimate and the Closing Disclosure. To further encourage mortgage shopping, the CFPB has also launched various tools and resources that help consumers make more informed decisions during the mortgage searching process. See http://www.consumerfinance.gov/owning-a-home/ for more details.

[17]We recognize that some borrowers will obtain financing from an institution that is not located in the same county as the purchased property. Since an institution that is present and active in a county will typically approve multiple mortgage applications in a single year, we do not include in some of our analysis institutions with a single

market of an individual with a given set of characteristics equals the total number of applications filed in that market by all individuals with this set of characteristics divided by the number of loan originations taken out by these individuals. To preview our estimation strategy, we first estimate how different characteristics affect search propensity and then use the composition of these characteristics to construct the baseline search intensity for markets.

The list of characteristics that we distinguish is dictated by the availability of data in the HMDA database. We thus group applications in a county that have the same combination of values for race, gender, ethnicity, loan purpose, owner occupancy, loan type, loan amount level, and income level (the last two are continuous variables and are discretized as noted below). Race includes Caucasian, Black or African American, Asian, and others.[18] Gender includes male, female, and not provided. Ethnicity includes Hispanic or Latino, not Hispanic or Latino, and not provided. Loan purpose includes home purchase, home improvement, and refinancing. Owner occupancy includes owner-occupied as a principal dwelling and not owner-occupied as a principal dwelling.[19] Loan type includes Conventional (any loan other than FHA, VA, FSA, or RHS loans), FHA (Federal Housing Administration)-insured, VA (Veterans Administration)-guaranteed, and FSA/RHS (Farm Service Agency or Rural Housing Service) loans. For loan amount, the four types we defined are the four quartiles within each state. Income types are defined in the same way, with an additional type containing those that did not report their income (around 5% of total observations in 2014). There are combinations of group characteristics with zero observations in some counties.

Table 1.2 reports the composition of individual/property/loan characteristics among the applications for all counties and those with no more than 300 originations in 2014, which we define as small markets, i.e., low population markets.[20] These markets, which for brevity we will refer to as "small markets", are of special importance because they are unlikely to be subdivided into distinct sub-markets (as big-city markets might be) and borrowers with the same observable

---

mortgage approval.

[18]"Others" includes individuals for whom the field is not provided and those who belong to the two demographic groups of Native American and Alaska Native, Native Hawaiian and Other Pacific Islander. These two demographic groups have too small a number of observations to include separately in the estimation, and they have a high correlation with the not provided group (0.82 and 0.77, respectively).

[19]Second homes, vacation homes, and rental properties are classified as not owner-occupied as a principal dwelling.

[20]The median number of originations in a county is equal to 303.

characteristics are likely more homogeneous within them. Indeed, small markets are not only less heterogenous but they also involve minimal aggregation, as the following series of statistics indicates. For all markets, there are 7,806 groups and 940,436 market-group level observations. The median number of groups per market is 162. For small markets, there are 3,476 groups and 128,045 market-group level observations, with the median number of groups per market being equal to 77. On average, there are 5.71 originations and 7.12 searches for every market-group level observation. But in small markets there are on average only 1.42 originations and 1.77 searches for each market-group level observation.

Mortgage applicants need not provide a response for some variables, which generates a "not provided" category. A discussion of this category may be useful. By far, the highest percentage of non-responses is for the race and ethnicity variables, where it is approximately a tenth of the sample. We suspect that a substantial number of individuals do not wish to classify themselves in one of the established categories. As an indication, 9.3% of individuals in the 2010 US Census described themselves as being part of two races or of "some other race". Others may prefer not to report a race on principle. For the purposes of our analysis, the group of individuals who do not report this information is treated as a distinct socio-economic group. More surprising is the fact that six percent of the applications do not list a gender. One can easily speculate on some possible reasons, but the "non-reported" category is also treated as a separate group in our analysis. Treating these individuals as distinct groups is reasonable if their behavior systematically differs from that of other groups, given that they are part of the market and their search activity has implications for the response of lenders and for the search activity of other borrowers.[21]

Our second major dataset contains all the fixed-rate conventional loans originated in the 50 states, Washington D.C., and Puerto Rico in 2014 (approximately 4.3 million individual loans) that are securitized by Fannie Mae, Freddie Mac, and Ginnie Mae.[22] In total, our dataset accounted for 73% of the first lien origination volume in 2014: 52% from originations securitized by the Government Sponsored Enterprises (Fannie Mae and Freddie Mac, referred to as the GSEs) and 21% from FHA/VA originations securitized by Ginnie Mae. The remainder 27% of first lien originations were not securitized by these entities and are not in our mortgage rate sample. We dropped

[21] We have estimated a few specifications after dropping these individuals from our sample, and obtained similar results.

[22] All HMDA applications in 2014 come from the same locations: the 50 states, Washington D.C., and Puerto Rico.

observations that have one or more than one of the following key variables missing: loan rate, credit score, LTV ratio (loan-to-value ratio), DTI ratio (debt-to-income ratio), loan purpose, loan amount, loan term, third party origination flag, number of borrowers, number of units, origination month, state, lender, and securitizer. The dataset also includes information on the occupation status and property type for GSE securitized loans and a first time buyer flag for 95% of the data. However, we do not have information on the points and fees borrowers pay. See Table 1.3 for a summary of these loan characteristics.

We also obtain county level demographic characteristics from the American Community Survey's 2014 5-year estimates, which contains information on population, education attainment, household income, per capita income, worker population, employment rate, labor force, occupied housing units, rent units, median gross rent, and median housing value. These variables are used as controls for market characteristics in our analysis. Key summary statistics are reported in Table 1.4 Panel A for all markets and small markets separately.

We constructed the Herfindahl-Hirschman Index (HHI) using the lenders' mortgage origination share in each county. This index is used as an indicator of each market's concentration level. To address possible endogeneity concerns with lender concentration at the county level, we also constructed two instruments for HHI. The first is the HHI calculated from 2007 HMDA data. The second instrument is the increase in HHI between 2007 and 2014 that is attributable to the merger and acquisition activities of banks and bank holding companies.[23] This activity is unlikely to be correlated with market-specific economic considerations, given that most banks operate over multiple markets. The M&A records are provided by the Federal Reserve Bank of Chicago and contain information that can be used to identify all bank and bank holding company acquisitions and mergers that have occurred since 1976.[24] Summary statistics on HHI and the associated instruments are reported in Table 1.4 Panel B.[25]

[23] More precisely, we compute the HHI using 2007 market share data after combining the market share of all banks and bank holding companies that have merged between 2007 and 2014. We then take the difference between that "counterfactual" HHI for 2007 and the actual HHI for 2007. This difference is used in our analysis as an instrument for the HHI in 2014. A similar instrument is used by Scharfstein and Sundaram (2014).

[24] The data files were obtained from https://www.chicagofed.org/banking/financial-institution-reports/merger-data. Updated versions of the data will be available from the National Information Center Bulk Data Download page.

[25] The Federal Deposit Insurance Corporation (FDIC) website has a listing of branch office locations and their annually reported deposits as of June 30, 2016. (Data is available back to 1994.) The listings provide branch office data by state, county, city and institution, downloadable at: https://www5.fdic.gov/sod/dynaDownload.asp?barItem=6.

1.3 Search Activity at the Demographic Group & Market Levels

Our first task is to pin down the differential search rates by borrower-type, where borrower-type consists of the intersections of the sets of borrower, property, and loan characteristics reported in the HMDA database. Recall that we do not observe individual borrower identifiers, and thus we do not know how many mortgage applications were initiated by each borrower of a given borrower-type. What we do observe is how many mortgage applications and mortgage originations were performed by all borrowers of a given borrower-type for properties in a given market. We know, then, the number of borrowers of borrower-type j in market m, which is equal to the number originations $Orig_{j,m}$ attributable to borrowers of that type in market m. We also know the total number of searches performed by all borrowers of each borrowing type in a market, which is equal to the number of applications approved by the lender or withdrawn by the borrower, $Apps_{j,m}$. From these, we compute the average number of "searches" by members of borrower-type j in market m, $Apps_{j,m}/Orig_{j,m}$. As noted earlier, we consider this ratio to be a proxy for less formal rate queries and other information acquisition efforts.

This ratio varies across markets on the basis of borrower-type and market characteristics. We postulate that the effects of these characteristics are additively or multiplicatively separable, and that unobserved factors that affect the search ratio are not systematically related to the observed ones. Moreover, the behavior of individuals of the same borrower-type is assumed to be the same across markets that are otherwise identical. This last condition is clearly a departure from reality, but there must be some substantial commonalities of behavior given the statistically significant findings.[26] These assumptions allow us to identify the differential search intensity associated with each borrower-type. In our framework, where the market characteristics are captured by a set of exhaustive dummies, borrower-type search propensity is identified from within market differences in per borrower search activity of each type.

To formalize, suppose we have K discrete borrower-type characteristics $T_1, T_2, \ldots, T_K$ and we denote the different values of T_k by $T_{k,1}, T_{k,2}, \ldots, T_{k,\kappa}, \ldots, T_{k,\kappa_k}$. Borrower-types, which to econ-

This database, which is not used in this version of the manuscript, may allow us to construct an alternative non-binary measure of lender presence in a county that does not directly depend on application volume.

[26] In fact, if each borrower-type is a mixture of underlying sub-groups with a composition that varies across markets, this would lead to an attenuation of our results.

omize on words we will often simply refer to as "groups", are defined by elements of the form $\{T_{1,l_1}, T_{2,l_2}, \cdots, T_{K,l_K}\}$, where $l_k \in \{1, 2, \ldots, \kappa_k\}$ and all possible combinations of $l_1 \times l_2 \times \cdots \times l_K$ are used. For example, a group would be defined with a specific combination of ethnic and socioeconomic (categorical) characteristics, applying for a mortgage of specific maturity and type, on a property within a given price range and characteristics. Many of these combinations, however, contain no individuals for at least some markets. Denote the value of characteristic k for group j by T_{k_j} and the proportion of group j in market m by $w_{j,m}$.[27] Out of the 6.7 million individual level applications, there were 940,436 unique group-market observations. There are some group-markets with only searches but no loan originations, likely because the borrower purchased a home in a different calendar year, or in a different market, or his purchase plans otherwise changed. This decreased the number group-market level observations used in our analysis to 812,446 for all markets.

However, it might be problematic to categorize large counties, some with more than 5,000 originations a year, as a single market. For example, Los Angeles county, with over 140,000 originations, can be reasonably divided into multiple overlapping markets. Moreover, the population in high population markets is more heterogeneous, even conditioning on observable characteristics, thus possibly attenuating the linkage between those characteristics and search propensity. Finally, mortgage lending is proportionately less important in large urban markets, and thus lenders may choose to stay in these markets for a variety of reasons besides mortgage lending. Therefore, the bulk of our analysis focuses on the bottom half of markets, i.e., those with no more than 300 originations. For these small markets, there were 227,894 individual level applications, which were bundled into 128,045 unique group-market observations. We do, though, report and discuss results based on the entire sample.

These group-market observations form our dataset for the first step of our analysis, which explains the variation in $Apps_{j,m}/Orig_{j,m}$ as a function of group and market characteristics. The independent variables are all binary indicator variables. Those where group characteristics take the value of one if that group has a particular attribute and the value of zero otherwise. Market characteristics, which are not of direct interest in this regression, are treated in the most flexible way and consist of an exhaustive set of market dummy variables. These dummies provide the

[27] These proportions are calculated using originated loans.

search activity of the omitted group in each market, and they additively scale the search activity of all groups in a market. We will refer to them as the market effects, or the adjusted-for-borrower-composition market level search activity. The regression we estimate is given by

$$\frac{Apps_{j,m}}{Orig_{j,m}} = \alpha_m + \sum_{k=1}^{K} \sum_{\kappa=2}^{\kappa_k} \beta_{k,\kappa} \mathbf{1}_{\{\kappa\}}(T_{k_j}) + \epsilon_{j,m} \tag{1.1}$$

where $\mathbf{1}_{\{\kappa\}}(T_{k_j}) = 1$ if $T_{k_j} = T_{k,\kappa}$ and 0 otherwise. The double sum consists of a series of dummies for each of the borrower-type characteristics, where one dummy per characteristic is dropped. Note that the independent variables take the exact same value for all the individual members of each group. Therefore, estimating this regression via weighted least squares, with weights equal to the number of individuals in a group, produces identical estimates to those we would have obtained using OLS had the individual-level data been available. A more efficient estimation approach is to estimate equation (1.1) via GLS, to account for the fact that idiosyncratic variability in search may systematically differ across group characteristics (including the size of the group).[28] GLS weights are almost linear in the number of individuals per group, so in this regard they do not depart much from analytic weights. But they do down-weight observations with attributes associated with high variance.

Because the group characteristics may affect search activity super-additively (possibly multiplicatively), we have also estimated the following log-linear specification of the above regression.

$$log\left(\frac{Apps_{j,m}}{Orig_{j,m}}\right) = \alpha_m + \sum_{k=1}^{K} \sum_{\kappa=2}^{\kappa_k} \beta_{k,\kappa} \mathbf{1}_{\{\kappa\}}(T_{k_j}) + \epsilon_{j,m} \tag{1.2}$$

Even though the results of regressions (1.1) and (1.2) are used primarily as an input to further analysis, we report them in Table 1.5. We observe some common features when comparing the linear and log specifications. For example, female borrowers search less than male borrowers, and both search less than those who do not report their gender. Hispanic borrowers search less than non-Hispanic borrowers. Borrowers refinancing their mortgage search more than borrowers

[28]GLS is implemented via iteratively re-weighted least squares as follows: 1. Estimate the unweighted linear regression and obtain the residuals. 2. Regress the residuals on the number of originations at the group-market level and on all controls. 3. Obtain the predicted value of residuals. 4. Rerun the original regression with weights proportional to the reciprocal of the squared predicted residuals. Repeat step 2 and step 3 until the estimated coefficients converge. In this study, we did 5 iterations.

purchasing a new property, not surprising given that they are under less time pressure. If a mortgage loan is VA-guaranteed, it's searched less when compared to other loan types.

The coefficients of relative search activity associated with a particular group characteristic are not of ultimate interest. What is of ultimate interest is whether the activity level in a market, adjusting for its group composition, is systematically related to borrower search propensity in a way that suggests spillovers or free-riding. To make this more concrete, suppose that non-Hispanic borrowers search more than Hispanic borrowers in the same market. We then ask the question whether borrowers in markets containing a high proportion of non-Hispanics search less than identical borrowers in markets containing a low proportion of non-Hispanics. This analysis could be performed by a "second-stage" regression, where the dependent variable are the estimates of the market dummies, α_m, from equations (1.1) or (1.2) and explanatory variables are the proportion of borrowers with each of the characteristics on the right hand side of those equations, plus any other market-level characteristics that can impact search activity. We would then compare the parameter estimates in "first-stage" equations (1.1) and (1.2) with those of this "second-stage" regression. We would expect that if a value of an attribute is associated with reduced search in the first-stage equations, the corresponding population weight coefficient in the "second-stage" would be positive.

However, this approach is fraught with two difficulties. First, there's too many characteristic coefficients to compare and the comparisons are not straightforward, e.g., a simple comparison of signs will not work because it is not invariant to the identity of the excluded category for each attribute. Second, some demographics may be proxies for other unobserved factors that affect search. Using a summary measure that combines all the estimates of the equations (1.1) and (1.2) eliminates the first difficulty. It also reduces the second, since it is unlikely that biases arising from the proxy effect of a demographic characteristic all point in the same direction. In particular, we use the group characteristics coefficients $\beta_{k,\kappa}$ and group market weights $w_{j,m}$ to construct every market's baseline search intensity BSI_m as:

$$BSI_m = \sum_{j=1}^{J} w_{j,m} \sum_{k=1}^{K} \sum_{\kappa=2}^{\kappa_k} \beta_{k,\kappa} \mathbf{1}_{\{\kappa\}}(T_{k_j}) \tag{1.3}$$

This index takes higher values for markets where borrowers have attribute values that are associated with higher market-adjusted search activity relative to borrowers with other attribute values.[29] Changing the identity of the omitted categories in equations (1.1) and (1.2) affects this index by a constant, i.e., it does not have an impact on the difference of this index across markets. When there are no interactions between attributes, as is the case in the analysis we report here, this index is identical to that obtained by multiplying the fraction of the borrowers that have a specific attribute value by the coefficient for that attribute value and summing over all attributes.

Before we proceed to the use of BSI_m to measure search spillovers, it may be useful to assess its correlation with various market characteristics. This would provide a rough measure about the "geography" of search propensity. Probably foremost among market characteristics is whether the market is rural or part of a metropolitan statistical areas (MSA). In our dataset, there are 1,587 counties, 245 of which are in an MSA. For both the linear and log model, BSI_m is on average smaller when market m is in a certain MSA, as shown in Table 1.6. Moreover, using county population/housing unit density data from the 2010 Census, we found that BSI_m decreases with both market m's population and housing unit density.

We now turn to estimation of the search externalities. We estimate the equation:

$$\hat{\alpha}_m = a + bBSI_m + cX_m + u_m \tag{1.4}$$

where $\hat{\alpha}_m$ are the estimated market effects and X_m are other market characteristics that may associated with differential search levels. The effect of these characteristics may not be causal; rather, they may be proxies for causal factors. For example, causal factors that may affect search could be the density of lender branches, the traffic conditions that permit visiting those branches, or the competitiveness of the local lending services. Though we do not have data on many such factors, they are likely to be related to key economic and demographic characteristics in the market. The full set of characteristics in X_m includes HHI, population, per capita income, number of owner occupied units, number of rent units, median rent, median housing value, percentage of population with a bachelor's degree, percentage of population who are minorities, percentage of

[29]The search levels of some groups may be more responsive to the aggregate market search propensity, i.e., the assumption that market-level search activity is additively or multiplicatively separable from the composition of borrowers may fail. For example, some groups may be more sensitive to changes in the price distribution. This would create a potential bias in the coefficients of equation 1.1 and 1.2. But the value of the BSI_m may be less biased if group-specific biases cancel out.

population 16 years and over that are in the labor force (willing and able to work but not necessarily employed), percentage of population 16 years and over that are employed, and the number of people working in this market (possibly commuting from other geographic locations) divided by the local population. Some of these characteristics enter in some form in the construction of the BSI_m (though for the BSI_m, the values of the characteristics correspond to the individuals who have obtained mortgages and not to the general population). These characteristics are the per capita income and the percentage of the population who are minorities. For robustness, we construct a partial set of market characteristics that excludes these variables. As we mentioned in section 1.2, the current value of HHI is instrumented with the 2007 value and the change in HHI driven by bank mergers and acquisitions since 2007. Observe that the dependent variable in equation (1.4) is an estimated parameter and thus contains sampling error. The standard error of the parameter, $\sigma_{\hat{\alpha}_m}$, is a measure of the sampling variability. Therefore, we first estimate equation (1.4) using standard errors of the market fixed effects as weights. We then estimate this regression using GLS, implemented via iteratively re-weighted least squares, where the error variance associated with an observation is a function of $\sigma_{\hat{\alpha}_m}$ and of all the independent variables.

The results are shown in Tables 1.7-1.8 for two versions of BSI. The former $BSI(AW)$ is constructed using the coefficients of relative search activity obtained from the analytic weights first stage; the latter $BSI(GLS)$ is constructed using the coefficients from the GLS first stage. We focus on the coefficient for BSI, b, which we interpret as reflecting the degree to which changes in the search propensity of the borrowers active in a market crowds out realized search activity. If there was no such crowd out, the estimate of b would have been zero: the search activity of borrowers of different characteristics would depend on market characteristics but not on the composition of borrowers in a market. A value of b in the $(-1, 0)$ range implies that changing the composition of borrowers so that, holding everything else fixed, search activity increases by one unit, would in fact lead to a partial compensatory reduction in search activity by all borrowers in a market: overall search activity would increase, but not one-for-one. Finally, if $b < -1$, then overall search level in that market decreases. In our results, $b \in (-1, 0)$ for $BSI(AW)$, which suggests that there's more applications in markets with high search intensity individuals. For $BSI(GLS)$, $b < -1$ in most cases, suggesting that crowd out is more than one-for-one and there's actually less applications in markets with high search intensity individuals.

Before turning our attention to the link between search propensity and prices, it is worthwhile re-emphasizing that the analysis of this section looks at a narrow definition of actual searches: filed applications. These could be a proxy for less involved search activity that does not yield a binding mortgage offer by a lender. In that case, the estimate of b would be interpreted in the same way, if formal applications are more or less linearly related with informal inquiries. Word-of-mouth tips, however, may also serve the same role as "searches". If high search propensity individuals are also prone to asking individuals from their social circle for mortgage related information, this can substitute for formal searching as well. In that case, even if the observed coefficient b is algebraically smaller than -1 (absolute value greater than 1), the total amount of actual search activity in a market may increase with *BSI*. Search activity, including information exchange among borrowers' social circles, would leave a "signature" in the transaction prices. We turn to this next.

1.4 Search Propensity and Quote vs Transaction Prices

Search activity and information spillovers between consumers influence the distribution of transaction prices, for any given distribution of posted prices. We proceed to measure the extent to which this is the case in the mortgage market, and draw inferences on the nature of search externalities. A major obstacle is that the HMDA database does not contain the rates offered by the financial institution. These must be obtained from other sources and combined with our HMDA sample. Unfortunately, there exists no publicly available mortgage rate data.[30] Therefore, we must "back-out" prices from mortgage transactions securitized by Fannie Mae, Freddie Mac, and Ginnie Mae.

Our starting observation is that in the United States, mortgage pricing is typically uniform within a state, because financial institutions fear possible exposure to allegations of "redlining". This observation has the following three implications. First, in "backing-out" firm pricing from transaction data, we can pool together all transactions involving a lender to the state level. Second, the search intensity of borrowers in a market has an attenuated effect on the prices a lender charges. These prices would reflect the competitive conditions in all the markets that a lender

[30] Alexandrov and Koulayev (2017) use a proprietary dataset collected by the CFPB.

operates in; for most lenders each individual county is a small component of their total market (something that is particularly true for counties that are "small" markets). Therefore, third, search activity in a market will affect the distribution of prices available to borrowers in that market primarily through its impact on the presence of financial institutions, i.e., through the entry decision rather than through the pricing decision. Search activity will affect the quotes a borrower actually receives through the sampling probability of each lender. It will also affect the probability that a borrower accepts any given offered rate, i.e., it will affect the distribution of transaction prices given the distribution of price offers. In this section, we focus on the computation and comparison of these price distributions, while in the next section we look into the locational decisions of financial institutions.

In using the mortgage transaction data to back-out prices, we note that financial institutions offer interest rates based on information that is available on the mortgage application. All key items in that mortgage application are available to us. However, lenders typically offer to each borrower the opportunity to trade-off a lower interest rate with a higher upfront payment, known in the industry as points. The points chosen by each borrower are not available to us. Thus, a low observed interest rate in our transaction data may reflect a high points payment, and vice versa. If borrowers with the same observable characteristics systematically chose different points depending on the financial institution that they transact with, this would render it impossible to ascertain which institutions are more expensive than the others. In what follows, we assume that the typical choice of points does not vary across institutions, conditional on borrower-type, and that the trade-off between points and interest rates is the same across institutions, i.e., institutions may vary in the rates they charge, but not in how rate discounts relate to points paid.

If these conditions are approximately met, we then compute the adjusted-price of an institution, which we will often simply refer to as price, from a regression of the transaction interest rate for a loan in the Fannie Mae, Freddie Mac, and Ginnie Mae database on borrower/loan characteristics and lender-state fixed effects. Formally, we estimate the equation

$$P_{i,l,s} = \mu_{l,s} + \zeta Z_i + e_{i,l,s} \tag{1.5}$$

where $P_{i,l,s}$ is the rate paid by borrower i to lender l in state s, $\mu_{l,s}$ are lender-state fixed effects, and

Z_i is the full set of rate-relevant characteristics, including credit score, loan-to-value ratio, debt-to-income ratio, and others.[31] The fixed effects capture the pricing of each lender in a state, after controlling for all other observable characteristics that might affect loan rates. The specification of equation (1.5) obtains an average measure of priciness of a financial institution for all borrower-types within each state. It is possible, in fact likely, that some institutions offer competitive rates for some types of borrowers (say those with high loan-to-value ratios) while other institutions offer competitive rates for other types of borrowers. This suggests that the lender dummies could also be interacted with some key borrower characteristics. For example, lenders who offer low interest rates for FHA loans may offer higher rates for non-FHA loans, and vice versa. However, we believe that the simple lender fixed effects provide an adequate measure of price dispersion in a market, even though the heterogeneity of pricing would be relevant for interpreting our results.[32]

The results of the price regression are reported in Table 1.9. Except for the lender fixed effects, the regression coefficients are not further utilized in this chapter. However, it is worth pointing out that the estimates are as expected, which is somewhat reassuring about our conjecture that the use of points is not strongly correlated with borrower characteristics (and hopefully, then, not strongly correlated with the financial institution). The characteristics associated with lower rates are, a better credit score, a larger loan, a mortgage associated with a home purchase, especially for a single-unit purchase. The characteristics associated with higher rates are high LTV ratios and DTI ratios, longer loan terms, and a retail transaction.

The estimates of the lender fixed effects are used to construct the price distribution in each market as follows. We first match the lenders in the price database with HMDA's respondent ID.[33] This matching is crucial because county-level transaction volume information is only available in HMDA. We next compute, based on each lender's fixed effect, the lender's rate for a "typical" borrower, fixing the borrower characteristics to average national values. We refer to this rate, $p_{l,s} = \mu_{l,s} + \zeta \bar{Z}$, as the lender's adjusted price, or simply price. We then use the institutional fact that

[31] We also created dummies for LTV and credit score brackets according to Fannie Mae's 2017 LLPA matrix and ran the same linear pricing equation with discrete notches instead of continuous LTV and credit score variables. The rankings of lender expensiveness had a 99.67% correlation and the predicted lender prices had a 99.73% correlation with our current model.

[32] See Chapter 2 for more detailed discussions on the heterogeneity of mortgage pricing.

[33] We were able to match the lenders for 97% of our Fannie, Freddie, and Ginnie loans.

lenders' mortgage pricing is uniform within states to obtain the price distribution in each county. We first construct the raw distribution of (adjusted) price in market m by computing the fraction of lenders active in a market with $p_{l,s}$ below any value x. In calculating this raw distribution, which we denote by $F_m^L(x)$, each lender with at least one mortgage approval in a county gets the same weight. If prospective borrowers were equally likely to obtain a quote from any of the lenders operating in a county, then this would be the relevant price distribution they would face. However, quite clearly some institutions are more likely to receive rate inquiries than others, if only because these institutions differ in size. A better approximation for the price distribution that consumers are facing could be obtained by weighting each lender's price by its number of branches in the county. However, branch information is only available for members of the FDIC, which excludes all non-bank issuers who are a major part of this market. Other measures of lender size, such as deposits, would be subject to the same limitation. Identifying additional measures of physical presence would still omit lenders operating via the Internet. Therefore, the distributions that can be constructed which would apply to all lenders are those based on equally weighted (i.e., unweighted) lender prices, or those weighted by information available in the mortgage database.[34]

Beyond the unweighted price distribution, we have computed two weighted price distributions. The first is based on the number of mortgage applications filed to lenders (and not rejected by them), which we refer to as the offer or quote price distribution. Let the number of applications to lender l in market m be $Apps_{l,m}$, and a lender's market share of applications by $s_{l,m}^A$. Then, the quote distribution, $F_m^A(x)$, is obtained by computing the fraction of market share weighted lenders with $p_{l,s}$ below any value x. The second weighted price distribution is obtained by using originations, $Orig_{l,m}$, to compute a lender's share of originations, $s_{l,m}^O$. This is used in the same manner as $s_{l,m}^A$ to obtain the transaction price distribution, $F_m^O(x)$.

Table 1.10 provides summary statistics for all three price distributions. We highlight two statistics. First, the quote mean is higher than the transaction mean, but only by a small amount. High priced lenders should have a smaller share in originations compared to their share of quotes, since borrowers would naturally choose the cheaper provider if they obtain more than one quote.

[34]The equal weighing of lenders will be less of a problem for small markets, since conditional on presence, there will be a smaller heterogeneity in the extent of that presence.

Therefore, a transaction weighted average price should be lower than the quote weighted average price.[35] A relatively small difference between the two would be expected since 80 percent of quotes end in originations. However, the difference is too small to explain by this high conversion percentage. A possible explanation is that, as mentioned earlier, changing the value of a borrower characteristic does not lead to the same price change for all lenders. As a result, one lender may be cheaper for a particular borrower than another lender, but that other lender may be more competitive than the first one for a different borrower. Therefore, search may lead borrowers to turn down some lenders for others, but the flows may partially cancel each other out. In other words, even though one lender may be generally more expensive than another lender, it does not necessarily follow that every borrower who gets a quote from both lenders will chose the latter over the former.

The second statistic that we want to highlight is that the difference between the bottom and top decile is approximately 0.4% for the lender average price distribution and 0.3% for the quote and transaction distributions. In other words, there is substantial disparity in the average price of lenders operating in a market. This compares with the 0.5% dispersion between the highest and lowest quote reported in Alexandrov and Koulayev (2017). If using the difference between the top and bottom deciles, their corresponding difference would have been 0.4% as well, despite the fact that we arrived at it through different methodologies. We used the full set of lenders, backed-out their pricing from transaction data, and weighted them by market share, but we did not have the rate sheets in our disposal and were not able to account for points paid. This consistency is certainly reassuring.

We now examine how these price distributions vary across markets, and most importantly, whether they are a function of the market's baseline search propensity measure that we constructed. Since lender mortgage pricing is uniform within states, any difference in the price distributions within states will be driven by the composition of active lenders in each market and differences in shares across markets. Holding the composition of active lenders constant, a market will have lower expected price quotes if borrowers in that market are more likely to file applications with lower price lenders. For each borrower, the transaction price equals the lowest of the quotes received. Therefore, the transaction price distribution will differ from the quote

[35]Following the same logic, the average lender price mean is higher than the quote mean.

price distribution to the extent that the lowest quote differs from the average quote; this gap will be larger if the typical borrower obtains more quotes or if borrowers who obtain the same number of quotes, obtain the second one from a particularly low priced lender.

The limitations on the unweighted average price in a market not-withstanding, investigating how this price depends on county characteristics and search propensity is a useful point of departure. Thus, the first equation we estimate is

$$E[p_{l,m}^{lender}] = a^l + b^l BSI_m + c^l X_m + u_m \tag{1.6}$$

where the expectation is the sample average of lender prices for the lenders that are active in market m (i.e., it is with respect to the distribution $F_m^L(x)$), and X_m is a vector of market characteristics that might be associated with market conditions that affect lender presence in a market (including characteristics that may affect search rates). We have also used the price means based on weighted data. Using application weights, we computed the mean price quote in market m, and estimate

$$E[p_{l,m}^{quote}] = a^q + b^q BSI_m + c^q X_m + u_m \tag{1.7}$$

where the price expectation is taken using quote share weights (i.e., it is with respect to the distribution $F_m^A(x)$) and X_m is a vector of market characteristics that might be associated with lender presence and differential search rates. We have also estimated this equation using $E[p_{l,m}^{orig}]$, the price expectation taken using the origination share weights, as the dependent variable. This estimation uses the number of originations in a market as weights. All three regressions have also been re-estimated using the value of the *BSI* obtained from the log-linear equation (1.2). We have also estimated the model with instruments for HHI, as in the estimation of the equation (1.4).

The results are shown in Tables 1.11-1.13 for the linear and log models. The average lender price declines with the baseline search intensity. The effect is larger and consistently statistically significant when the GLS-derived value of BSI is used as a measure of search propensity, and when the expanded set of market characteristics is used as regressors. The typical lender operating in a high search market is a low-price one. But these estimates are more tenuous when the value of *BSI* is derived from the analytic-weight regressions. The results are far stronger when moving to applications and transaction prices. The average quote and average origination price in a market strongly decline with either the GLS based or analytic-weight based *BSI*, with the effects being stronger for the former. This holds true for both the linear and log model, although the effects are more pronounced for the log model. Furthermore, for both models, the magnitude of decrease is somewhat larger for the origination price mean. This difference is small but statistically significant.[36]

A comparison of the results with respect to the sensitivity of prices to search propensity sheds light on the impact of search. The reduction of the quote price mean associated with *BSI* has two components. The first is that financial institutions are less likely to operate in markets with consumers who are prone to collect more information; the second is that among the institutions operating in those markets, informed prospective borrowers are more likely to file applications with cheaper ones. Recall that when we use the average price of a lender in a market, $E[p_{l,m}^{lender}]$, as the dependent variable in equation (1.6), the decline of that price with *BSI*, which gives the first of these two components, is small. Therefore, the driver is primarily differences in the probabilities with which borrowers file applications to lenders in different parts of the price distribution. This seems to suggest that borrowers rely on prior information from sources in their social network (or other sources) when they submit applications and this information results in directed search focusing on lenders with lower mortgage rates.

The distribution of origination prices is even more responsive to *BSI*. One possibility for this responsiveness is that borrowers in markets with high search propensity file more applications, and as a result they obtain a better transaction rate holding the distribution of offers fixed. However, as discussed in the preceding section, there is an essentially complete "crowd out" of the higher propensity to file more applications. The observed reduction in origination rates, then, must be an outcome of directed search. Informed individuals, when they file a second application, do so for an institution that has very competitive pricing.

In the second set of equations we estimate, we move beyond the central tendency in prices, and look at how the entire offer and transaction price distribution depend on underlying search

[36]We have estimated regressions explaining the difference $E[p_{l,m}^{quote}] - E[p_{l,m}^{orig}]$ (results not reported for brevity). The coefficient of *BSI* in these regressions is positive and statistically significant, indicating an increasing gap between offers and transaction prices in high search environments.

propensities. In particular, we estimate the equation:

$$Q[p_{l,m}^{lender}|\tau] = a^l + b^l BSI_m + c^l X_m + u_m \tag{1.8}$$

where $Q[p_{l,m}^{lender}|\tau]$ is the τ^{th} percentile of the lender price distribution $F_m^L(x)$. These equations are not estimated via quantile methods, where observations from all quantiles are used as the dependent variable and the quantile check function is used to re-weight the objective appropriately to yield parameter estimates for the desired value of τ. Rather, the dependent variable is directly the τ quantile of the distribution and the equation is estimated via linear regression, again weighted by the number of originations in the market. We also estimate equations using the application weighted prices:

$$Q[p_{l,m}^{quote}|\tau] = a^q + b^q BSI_m + c^q X_m + u_m \tag{1.9}$$

where $Q[p_{l,m}^{quote}|\tau]$ is the τ^{th} percentile of the quote distribution $F_m^A(x)$. Finally, we estimate specifications where the application weighted prices are replaced by the transaction weighted prices.

These regressions also have the same complement of explanatory variables as the mean regressions. However, in the results reported in Tables 1.14-1.16, the estimates of these other explanatory variables are omitted to conserve space. A higher value of *BSI* is associated with lower lender prices in most quantiles around the median and lower, but also at the topmost decile. The responses of application and originated weighted deciles to *BSI* are stronger, as expected from the mean regression results. More interestingly, these effects are generally larger and more significant for deciles around and below the 70th percentile. By and large the high-end of mortgage offers seems to be the same in all locations. But below that high-end, offers seem more likely to be low in high information environments, an effect that is even more pronounced for originations.

This latter effect can arise because mid and low price lenders are more frequent participants in high information markets, or it may arise because they receive a disproportionate number of applications in those markets. Alternatively, perhaps the participation of low priced lenders is not very responsive to the search/information in a market, but higher priced lenders are less likely to participate, skewing the distribution towards lower prices. These questions can be answered by looking at the participation margin of lenders in markets, which we take up in the next section.

1.5 Lenders' Market Presence

A financial institution's deposit rates and rates for other non-mortgage financial products often differ across the localities it operates in. However, the mortgage rates of a given lender tend not to vary within states, and sometimes do not vary across states either. As noted earlier, the main reason is the fear of being accused of "redlining". Therefore, lenders cannot tailor mortgage rates to the local market conditions, including adjusting them in response to the search intensity in a particular market. Rather, mortgage rates reflect the competitiveness in the entire set of markets in which lenders operate and also reflect their operating costs and brand name. Large lenders are generally more prominent, generate more traffic and could charge higher rates.

Whatever the optimal rate of a lending firm is, it can decide to be active or inactive in a particular market, i.e., it can decide whether to enter a market or stay out. A high search propensity area would be one where a lender can obtain smaller profits, all things equal. This would be particularly relevant for high price (high cost) lenders. They are the ones that increased search activity or increased information would hurt disproportionately, as borrowers would be able to identify and choose lower cost alternatives. Therefore, high price lenders should be particularly inclined to stay out of high search areas, leading to lower prices but due to selection. However, mortgages are only part of the sales portfolio of lenders. Search induced competitiveness in the mortgage market would have strong effects only if it is correlated with competitiveness for other products (other loans, deposit accounts, etc.). This is a strong, but plausible premise. Consumers who search intensively for mortgages and exchange information about their experience with each other are also prone to do so for other financial products.

With the above discussion in mind, we investigate whether lenders' entry decision into a certain market is affected by the market's baseline search intensity and the lender's state-wide mortgage rate. The lender's interest rate is (mostly) an exogenous object, determined primarily by factors other than the participation decision in that specific market. This is particularly true for small markets, since a lender's presence there is of very small importance and unlikely to affect state-wide pricing decisions. Recall that *BSI* is also an exogenous object: it is the propensity to search and obtain information and not the actual search in a market. We start by creating a lender-market level dataset as follows: 1) For each state s, we record all the lenders l active, i.e. lenders that have at least one search record from state s in the HMDA data.[37] 2) For each active lender l, if it has at least one search record in market m (that belongs to state s), we assign the entry indicator $E_{l,m} = 1$; otherwise we assign $E_{l,m} = 0$. Let the total number of lenders that are active in state s be L_s, and let the total number of markets in that state be M_s. Then, in this dataset, there are $L_s \times M_s$ observations for state s. Approximately 15% of the lender-market level observations in this constructed dataset have $E_{l,m} = 1$. We then match the lenders in each state with their price, $p_{l,s}$, as obtained from the estimation of equation (1.5).[38]

We use this dataset to estimate, via a probit regression, the probability that a lender is active in market m as a function of that market's BSI, the lender's price, and a host of other market characteristics. Because we want the marginal effect of BSI to vary flexibly with the lender's price, the specification we employ is a flexible spline with respect to $p_{l,s}$ and its interaction with BSI. In particular, we estimate:

$$\begin{aligned} Pr(E_{l,m} = 1) &= \alpha + \beta BSI_m + \delta p_{l,s_m} + \gamma BSI_m \cdot p_{l,s_m} + \sum_{\lambda=1}^{\lambda=\Lambda} \delta_\lambda \max(0, p_{l,s_m} - p_\lambda) \\ &\quad + \sum_{\lambda=1}^{\lambda=\Lambda} \gamma_\lambda BSI_m \cdot \max(0, p_{l,s_m} - p_\lambda) + \phi X_m + \nu_{l,m} \end{aligned} \tag{1.10}$$

where $p_\lambda (\lambda = 1, 2, ..., \Lambda)$ are knots for the spline function, the subscript s_m refers to the state s where market m is located in, and X_m are market characteristics among those used earlier in this chapter for market-level analysis.[39] The set of knots contains the 10th, 25th, 50th, 75th, and 90th percentiles of the rates in this lender-market level dataset.[40] Besides controlling for market-level characteristics, we also include state fixed effects.[41]

[37] We also tried a looser criterion of entry using all applications instead of only "searches" (as defined in the data section) as an indicator of being active in a certain market. The results are similar, although the *BSI* effects on entry are weaker.

[38] Since we only have price data for loans that are securitized by the GSEs or Ginnie Mae, roughly 60% of the actual lender entries ($E_{l,m} = 1$) in HMDA are matched to the corresponding lender prices. These lenders account for 78% of the originations in HMDA, and 97% of the mortgages in the GSEs and Ginnie Mae database.

[39] Note that for every lender p_{l,s_m} is actually the same for all markets in the same state.

[40] Another set of knots we tried contains all 9 deciles of the rates in this lender-market level dataset. The graphs associated with those regression results are much spikier, but the general tenor of the findings remains the same.

[41] The lender price is the only lender characteristic, because we want price to be the summary measure of all lender characteristics that affect its pricing structure. Ideally, we would like to have the expected price of the lender, to remove any small level of endogeneity between a lender's price and the entry decision, but we do not have enough lender characteristics to get reasonable estimates of it. Adding lender characteristics would remove some of the exogenous components of the observed price.

The results of this analysis are reported in Table 1.17. More instructive than the values of individual parameter estimates is the impact of *BSI* on the probability that a lender charging a particular rate is active in market m. In Figure 1.1 Panels A and B, we plot the derivative of the entry probability with respect to *BSI* as a function of a lender's lending rate for the linear and log model, respectively. Because the cumulative probability function is a non-linear function of the linear index of the probit regression, the value of the derivative will depend on the value of *BSI* at which it is evaluated. Therefore, these figures contain a number of different piecewise linear curves for different values of *BSI*. We choose to plot the function for the three quartiles, plus two extreme deciles. Since all lines are very close to each other, obtaining a single line for the marginal effect of *BSI* on entry probabilities would be easier to read.

To do so, we first calculate the derivative of entry probability with respect to *BSI* for every lender-market observation using the estimated coefficients of equation (1.10) and denote these predicted derivatives as $\frac{dP(\hat{E}_{l,m})}{dBSI_m}$. We then run a spline regression with these (predicted) derivatives as the dependent variable and the lender's interest rate as the independent variable. We use the 9 deciles of interest rates as knots. This is essentially a semi-parametric smoothing of the predicted values of the derivatives against the lender's interest rate. The resulting function and bootstrapped confidence intervals are plotted in Figure 1.1 Panels C-F, for $BSI(GLS)$ and $BSI(AW)$, linear and log model respectively. We note that the derivative of the entry probability with respect to *BSI* starts off positive when rates are lower than 4.1%, indicating a higher *BSI* increases the entry probability of cheaper lenders. When rates surpasses 4.1%, the derivative becomes negative, signaling a higher *BSI* deterring the entry of more expensive lenders. As we expect, the negative marginal effect increases as rates increase, up to the point where rates equal 4.4%. Afterwards, the marginal effect reduces, while still being negative. The most expensive of lenders in a market probably obtain much of their revenue from unrelated sources. Though their presence in a market is adversely affected by *BSI*, the negative effect is not as strong as those lenders that are just a bit less costly than them.

To summarize, these results show that markets with high search and information acquisition propensity, not only have lenders that are offering lower rates but also have fewer active lenders. This latter effect is primarily concentrated on the mid and mid-high priced lenders. Only the lenders with particularly low rates are attracted to these markets, and the effect is not large.

High levels of consumer information reduces search frictions and may leave lenders with less profitable opportunities.

We next use this lender dataset to look at how applications and originations at the lender level depend on the value of the search propensity in a market. Considering, as above, the full set of lenders that are active in a given state, we estimate a zero-inflated poisson model, with the values of applications or originations for a particular lender in a market as the dependent variable. The zero-inflated poisson model is used because of the prevalence of zeros for lenders that did not enter a specific market.[42] We use a logit model to characterize the excess zeros, which yields the log likelihood $\ln(L)$ function

$$\begin{aligned}\ln(L) &= \sum_{j\in S} ln\Big[F(\boldsymbol{z_j\gamma}) + \{1 - F(\boldsymbol{z_j\gamma})\}\cdot \exp(-\lambda_j)\Big] \\ &+ \sum_{j\notin S}\Big[ln\{1 - F(\boldsymbol{z_j\gamma})\} - \lambda_j + \boldsymbol{x_j\beta}\cdot y_j - ln(y_j!)\Big]\end{aligned} \tag{1.11}$$

where F is the inverse of the logit link, and S is the set of observations for which the number of applications/originations y_j is zero. In the zero-inflated poisson analysis, we use the same set of independent variables as in equation (1.10), i.e. $\boldsymbol{z_j} = \boldsymbol{x_j}$.[43]

The estimation results are shown in Table 1.18. As in the analysis of lender entry, it is more instructive to use the estimates to calculate how the number of lenders' applications and originations in a market is affected by the baseline search intensity. We follow a similar approach to that we used in the entry analysis. For each lender, we compute the derivative of predicted originations with respect to *BSI* in its market. We then use this predicted derivative as the dependent variable in a spline regression against the lender's interest rate, with knots at the 9 deciles of the price distribution. The resulting graph of *BSI*'s marginal effect on the number of originations for lenders is shown in Figure 1.2.[44] The derivative is larger than zero when the rate is smaller than 4.17%, indicating the number of predicted originations increasing with *BSI* for low cost lenders.

[42]The zero-inflated poisson model naturally accounts for the non-negative nature of the dependent variable. It also contains implicit interactions between the effects of various regressors which is appropriate in this case. According to the Vuong test, a zero-inflated poisson is statistically more appropriate for our data than the standard poisson model.

[43]We obtain very similar results when using the partial set of market controls in the zero-inflated poisson model.

[44]We show the derivative of predicted originations with respect to *BSI*, the derivative of predicted applications with respect to *BSI* looks very similar and hence is omitted for brevity.

At rates higher than 4.17%, the derivative becomes negative and continues to decrease. The minimum level is reached when the rate is around 4.3%; thereafter, the derivative starts to increase but remains negative. As is the case with lenders' choice of entry, originations for lenders at the highest price levels are less affected by information-induced competition than their counterparts that are a little cheaper.

One difference between the entry and origination plots is that originations of low price lenders respond very strongly to baseline search activity. This reflects the fact that low price institutions are not only (marginally) more likely to enter a high search market, but make a disproportionate amount of originations. This intensive margin is driven by the fact that in these high search/high information markets, borrowers are more likely to identify low price lenders and obtain loans from them.

1.6 Additional Analysis

The results discussed in the preceding sections are rather comprehensive. We have estimated additional specifications, for the purpose of assessing the stability of our findings. The associated results are largely qualitatively similar to those reported, though differences are quantitatively material in some instances. The findings of additional analysis are described here.

1.6.1 Additional Results Using Small Markets

The specifications reported in this chapter vary in a number of dimensions with respect to the estimation method. The group-level search intensity regressions can be estimated either via analytic weights or via GLS (Generalized Least Squares). Both are reported. The market level search activity regressions can be estimated either via variance weighted least squares or GLS. In either variant HHI can be instrumented, the set of explanatory regressions can vary in comprehensiveness and include either version of the baseline search intensity index. We have estimated all the combinations of the possible specifications that are not reported here. Generally, the coefficient of *BSI* is larger (in absolute value) when this variable is constructed based on the GLS group-level regressions. The coefficient of *BSI* tends to also be higher when the market-level regression itself is estimated via GLS. A larger regressor set also results in larger point estimates, but instrumenting for HHI leads to no material differences. None of these results, which are available upon request, contradict our conclusions with regards to substantial search externalities. Indeed, the "low end" of the point estimates for the coefficient of *BSI* is approximately -0.5, suggesting that 50 percent of an exogenous increase in search activity by some group will be compensated from reduction in the search of others.

The price regressions can also be estimated using either "version" of *BSI*, with origination (analytic) weights or via GLS, with long or short regressor lists, and with or without instrumentation for HHI. We have estimated all combinations not reported here. The general ranking of the unweighed price distribution based on lender counts being the least sensitive to search intensity and the distribution of origination rates being the most sensitive remains valid. The effects are also negative, as expected. There is somewhat smaller robustness in this result with respect to the count-based distribution of prices, which in the results we report has the smallest effects to begin with. Additional results for the lenders' entry decision have been briefly mentioned earlier. Generally, using the analytic weight derived value of *BSI* results in somewhat smaller sensitivity of entry activity to search intensity, but the general pattern is similar.

1.6.2 Analysis Based on All Markets/Counties

The reported results all focus on small markets, i.e., counties with a relatively small number of borrowers (and a small number of lenders). Counties with a large population and many lenders effectively constitute multiple markets and also involve more complicated firm presence decisions. For example, a lender may enter in a small part of Los Angeles county, but this does not imply that this lender is an effective choice for borrowers in every part of that county. Nor does it imply that the lender's entry decision was largely driven by the retail (individual and small business) market segments, as financial institutions find that maintaining presence in major cities is important for other reasons. Moreover, the scale of lender entry, which we largely side-step in our analysis, increases in relevance in large markets. In small markets, there is less heterogeneity in the extent of lender presence; even large banks will only have relatively few branches.

Nonetheless, we have re-estimated the full set of regressions using observations from all 3,212 counties in our HMDA dataset. The results are available upon request, but we briefly summarize them here. With regards to the group-level search intensity regressions, the relative ranking of the effect of borrower characteristics on search is the same as that for small markets for the log regression, but differs somewhat for the linear model. More specifically, Caucasian and African American, and Conventional and FHA-insured have reversed rankings. The estimated value of the *BSI* for the 1,234 counties that belong to an MSA tends to be smaller than that for the other counties. The market level search activity regressions yield smaller (in absolute value) coefficients for *BSI*, though these are still consistently negative and statistically significant. When the regression is estimated via GLS and the GLS-based *BSI* is used as a regressor, its coefficient is very close to minus one, indicating an approximately one-to-one crowd-out. When variance weights are used and the analytic weights based *BSI* is used as a regressor, the corresponding coefficient is a bit larger in absolute value. The pattern of quote and transaction price responses are mostly similar to that of small markets, albeit the *BSI* coefficient being generally smaller in absolute value. Minor discrepancies do occur when we use the partial control set (excluding per capita control and minority percentage), for some cases, the decrease in quote price is larger than that in transaction price. However, the difference is relatively small. The most important difference is when the price regressions are estimated with origination weights using the analytic weights based *BSI*. In that case, an increase in *BSI* often has no statistically significant effect on prices (and of the wrong sign for some quantiles). Generally, this attenuation of estimates is expected given that these markets are composites of constituent markets.

Lenders' market presence in all markets mostly resemble that in small markets. Some observations, obtained from the results using the GLS-based *BSI*, are worth pointing out. First, the entry probability derivative is now negative for all rates, signaling the fact that even for low cost lenders, a higher *BSI* decreases their probability of entry. The effect is still stronger for high priced lenders, indicating pricier lenders being more sensitive to information-induced market competitiveness than low priced lenders. Second, the derivative of entry probability with respect to *BSI* rebounds less for high priced lenders (those with rates above 4.4%), suggesting that their entry probability remains as sensitive to *BSI* as their slightly cheaper counterparts. Third, although the derivative of applications/originations with respect to *BSI* is on average much larger in absolute value, the rate where the derivative switches from positive to negative and the rate where the smallest derivative occurs are almost exactly the same as that for small markets.[45] Re-estimating the entry regressions using the analytic weights based *BSI* leads to the following differences relative to GLS-based *BSI* results. The (negative) entry probability derivative is larger in absolute value for rates lower than 4.1%, implying cheap lenders' probability of entry being more responsive to *BSI*. For the linear model, the derivative of applications/originations with respect to *BSI* is always negative, in contrast with previous cases where the derivative is positive for rates lower than 4.1%. Moreover, the range of the derivative functions are now smaller, signaling the number of applications/originations being less sensitive to *BSI*.

1.7 Concluding Remarks

The rate of approved mortgage applications to originations varies substantially across borrowers of different characteristics and across geographic locations. Our work shows that much of this variation is borrower's response to changes in the market environment induced by the search propensity of the typical borrower in the market. Markets where the typical borrower has a high search propensity and is thus typically better informed of mortgage rates (possibly from the experience of other borrowers) are markets where, all else being equal, fewer mortgage applications need to be filed. In part, this is because better informed borrowers target their applications to lenders who are likely to be a better fit for the borrowers' needs, along with competitive rates. In part, this is because high cost/high price lenders are less likely to enter high information markets, thus blunting incentives for search.

Though the existence of such search externalities has long been discussed in the theoretical and empirical literature, there has been little evidence to date that they are materially important for the determination of the market equilibrium. This chapter shows that they can, in fact, be quantitatively very important. An exogenous increase in search activity will be mostly compensated by a reduction in the search of other consumers. The reduction in equilibrium prices can plausibly explain only a fraction of this response; the rest is explained by an increase in search efficiency.

[45]When analyzing the number of applications/originations we used a zero-inflated negative binomial model instead of a zero-inflated poisson mode, because the former fits better. Nevertheless, the results are not affected much by this model specification change.

1.8 Tables and Figures

Table 1.1: **Actions Taken for HMDA Applications and Data Selection**

Panel A: Action Taken in HMDA Applications	Obs.	Percent
Loan originated	5,466,417	60.92%
Application approved but not accepted	322,802	3.60%
Application denied by financial institution	1,589,281	17.71%
Application withdrawn by applicant	1,045,079	11.65%
File closed for incompleteness	364,126	4.06%
Pre-approval denied by financial institution	123,077	1.37%
Pre-approval approved but not accepted (optional reporting)	62,966	0.70%
Total	8,973,748	100.00%

Panel B: HMDA Data Selection Criteria	Remaining Obs.
(1) Applications for first-lien, 1-4 family homes	8,973,748
(2) Drop applications not categorized as searches	6,834,298
(3) Drop if county code missing	6,787,902
(4) Drop if applicant is not a natural person	6,700,772
(5) After grouping by characteristics	940,436

Note: In Panel A, the applications we include are for first-lien, 1-4 family homes only.

Table 1.2: **Borrower/Loan Characteristics Distribution from HMDA**

	Small Markets		All Markets	
Race	Obs.	Percent	Obs.	Percent
Caucasian	199,267	87.44%	5,189,597	77.45%
Asian	1,073	0.47%	384,865	5.74%
Black or African American	9,531	4.18%	382,671	5.71%
American Indian or Alaska Native	2,453	1.08%	41,637	0.62%
Native Hawaiian or Other Pacific Islander	283	0.12%	26,771	0.40%
Not Provided	15,287	6.71%	675,231	10.08%
Gender				
Male	163,610	71.79%	4,488,120	66.98%
Female	53,550	23.50%	1,807,624	26.98%
Not Provided	10,734	4.71%	405,028	6.04%
Ethnicity				
Not Hispanic or Latino	194,449	85.32%	5,469,814	81.63%
Hispanic or Latino	17,963	7.88%	591,839	8.83%
Not Provided	15,482	6.79%	639,119	9.54%
Loan Purpose				
Home Purchase	110,736	48.59%	3,621,387	54.04%
Refinancing	104,955	46.05%	2,892,392	43.17%
Home Improvement	12,203	5.35%	186,993	2.79%
Owner-occupancy Status				
Owner-occupied	196,461	86.21%	5,951,496	88.82%
Not Owner-occupied	31,433	13.79%	749,276	11.18%
Loan Type				
Conventional	160,692	70.51%	4,953,252	73.92%
FHA-insured	28,791	12.63%	975,801	14.56%
VA-guaranteed	19,534	8.57%	611,602	9.13%
FSA/RHS	18,877	8.28%	160,117	2.39%
Total	227,894	100.00%	6,700,772	100.00%

	Mean	SD	Mean	SD
Loan Amount (in thousands)	124.93	131.82	228.24	231.65
Income (in thousands)	82.58	127.24	112.84	167.02

Note: Income statistics were computed using 6,366,046 observations, the remaining 334,726 observations have income missing.

Table 1.3: **Loan Characteristics from GSEs and Ginnie**

Panel A: Discrete Characteristics	Obs.	Percent
Mortgage Securitizer		
Fannie Mae	1,675,173	43.37%
Freddie Mac	1,078,121	27.91%
Ginnie Mae	1,109,448	28.72%
Third Party Origination Flag		
Broker	373,867	9.68%
Correspondent	1,360,707	35.23%
Retail	2,128,168	55.09%
Loan Purpose		
Purchase	2,356,485	61.01%
Non-purchase	1,506,257	38.99%
Number of Units		
1 Unit	3,781,789	97.90%
2 Units	58,194	1.51%
3 Units	12,117	0.31%
4 Units	10,642	0.28%
Number of Borrowers		
1 Borrower	2,034,853	52.68%
2 Borrowers	1,811,900	46.91%
$\geq$ 3 Borrowers	15,989	0.41%
Total	3,862,742	100.00%

Panel B: Discrete Characteristics (GSEs Only)	Obs.	Percent
Owner-occupancy Status		
Owner-occupied	2,362,188	85.79%
Not owner-occupied	391,106	14.21%
Property Type		
Single Family	1,786,628	64.89%
Non-Single Family	966,666	35.11%
Total	2,753,294	100.00%

Panel C: Continuous Characteristics	Mean	SD
Origination Rate (%)	4.29	0.46
Credit Score	729.13	56.02
Loan-to-value ratio	81.33	18.40
Debt-to-income ratio	35.49	9.94
Loan Amount (in thousands $)	200.64	112.21
Origination Term (in years)	27.06	5.95

1. In Panel A, we report the characteristics that the GSEs (Fannie Mae & Freddie Mac) and Ginnie Mae have in common. In Panel B, we report owner-occupancy status and property type, two characteristics that are exclusive to the GSEs.
2. All the continuous characteristics in Panel C are reported in both the GSEs and Ginnie Mae's data. The total number of observations we used to calculate the summary statistics is 3,862,742.

Table 1.4: **Market Characteristics Summary Statistics**

Panel A: American Community Survey (ACS) Variables

	Small Markets					All Markets				
	Min.	Max.	Mean	SD	Obs.	Min.	Max.	Mean	SD	Obs.
ln(Population)	6.05	11.20	9.21	0.85	1,586	6.05	16.12	10.28	1.45	3,211
ln(PercapitaIncome)	8.76	10.83	9.95	0.27	1,586	8.76	11.06	10.04	0.27	3,211
ln(OwnerUnits)	3.69	9.59	7.94	0.83	1,586	3.69	14.22	8.98	1.39	3,211
ln(RentUnits)	3.76	8.88	6.89	0.88	1,586	3.76	14.37	8.02	1.54	3,211
ln(MedianRent)	5.51	7.15	6.36	0.19	1,586	5.51	7.50	6.50	0.25	3,211
ln(MedianValue)	9.90	13.15	11.43	0.33	1,586	9.90	13.72	11.67	0.45	3,211
BachelorsPct	0.03	0.41	0.11	0.04	1,586	0.03	0.41	0.13	0.05	3,211
MinorityPct	0.00	0.96	0.17	0.19	1,586	0.00	0.96	0.17	0.17	3,211
WorkerPct	0.19	5.17	0.79	0.26	1,586	0.19	5.17	0.81	0.24	3,211
EmploymentPct	0.66	1.00	0.91	0.05	1,586	0.66	1.00	0.91	0.04	3,211
LaborForcePct	0.21	0.90	0.56	0.08	1,586	0.21	0.90	0.59	0.08	3,211

Panel B: Herfindahl-Hirschman Index (HHI)

	Small Markets					All Markets				
	Min.	Max.	Mean	SD	Obs.	Min.	Max.	Mean	SD	Obs.
HHI (2014)	0.02	1.00	0.14	0.12	1,587	0.01	1.00	0.10	0.10	3,211
HHI (2007)	0.00	1.00	0.11	0.09	1,587	0.00	1.00	0.08	0.07	3,206
HHI (Difference)	0.00	0.15	0.01	0.02	1,587	0.00	0.15	0.02	0.02	3,206

1. The American Community Survey provides data on 3,220 markets. Here we only summarize the 3,211 markets that are matched with the HMDA data and have at least 1 origination in 2014.
2. The variable "WorkerPct" is defined as the percentage of workers working in a specific market (possibly commuting from other markets) divided by the labor force in that market.
3. HHI is 0 when there are no originations in that market for the entire year.

Table 1.5: **Search and Group Characteristics**

	Analytic Weights				GLS			
	Linear Model		Log Model		Linear Model		Log Model	
Caucasian	0.076***	(0.003)	0.057***	(0.002)	0.061***	(0.003)	0.044***	(0.002)
African American	0.040***	(0.005)	0.025***	(0.004)	0.029***	(0.005)	0.020***	(0.003)
Asian	-0.030***	(0.009)	-0.026***	(0.006)	-0.006	(0.011)	-0.008	(0.007)
Male	-0.032***	(0.004)	-0.022***	(0.003)	-0.034***	(0.004)	-0.023***	(0.003)
Female	-0.068***	(0.004)	-0.051***	(0.003)	-0.061***	(0.004)	-0.042***	(0.003)
Hispanic	0.015***	(0.005)	0.011***	(0.004)	0.019***	(0.006)	0.011***	(0.004)
Non-hispanic	0.068***	(0.003)	0.052***	(0.002)	0.046***	(0.004)	0.031***	(0.003)
Purchase	0.064***	(0.003)	0.053***	(0.002)	0.053***	(0.003)	0.038***	(0.002)
Refinance	0.112***	(0.003)	0.087***	(0.002)	0.071***	(0.003)	0.048***	(0.002)
Owner-occupied	0.076***	(0.002)	0.060***	(0.002)	0.040***	(0.002)	0.027***	(0.002)
Conventional	0.041***	(0.004)	0.032***	(0.003)	0.025***	(0.004)	0.017***	(0.003)
FHA-insured	-0.001	(0.004)	-0.005	(0.003)	-0.005	(0.004)	-0.005	(0.003)
VA-guaranteed	-0.028***	(0.004)	-0.024***	(0.003)	-0.017***	(0.004)	-0.012***	(0.003)
Loan Bin 2	-0.019***	(0.002)	-0.019***	(0.002)	-0.010***	(0.002)	-0.007***	(0.002)
Loan Bin 3	-0.022***	(0.003)	-0.021***	(0.002)	-0.012***	(0.003)	-0.009***	(0.002)
Loan Bin 4	-0.012***	(0.003)	-0.012***	(0.002)	-0.002	(0.004)	-0.002	(0.003)
Income Bin 1	0.041***	(0.005)	0.033***	(0.003)	0.023***	(0.005)	0.015***	(0.003)
Income Bin 2	0.006	(0.005)	0.005*	(0.003)	-0.004	(0.004)	-0.004	(0.003)
Income Bin 3	-0.002	(0.005)	-0.001	(0.003)	-0.009**	(0.004)	-0.007**	(0.003)
Income Bin 4	-0.002	(0.005)	-0.001	(0.003)	-0.006	(0.005)	-0.005	(0.003)
Constant	0.867***	(0.007)	-0.108***	(0.005)	0.961***	(0.007)	-0.026***	(0.005)
Observations	108,098		108,098		108,098		108,098	
R^2	0.066		0.086		0.066		0.072	

Note: We use the number of underlying individual observations in each group as analytic weights.

Huber-White standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.6: *BSI* **Distribution**

Panel A: Analytic Weights	Min.	Max.	Mean	SD	Obs.
Linear Model					
BSI	0.134	0.351	0.267	0.020	1,587
BSI (MSA)	0.134	0.305	0.260	0.025	245
BSI (non-MSA)	0.145	0.351	0.268	0.019	1,342
Log Model					
BSI	0.105	0.273	0.209	0.016	1,587
BSI (MSA)	0.105	0.237	0.202	0.020	245
BSI (non-MSA)	0.110	0.273	0.210	0.015	1,342

Panel B: GLS	Min.	Max.	Mean	SD	Obs.
Linear Model					
BSI	0.059	0.211	0.163	0.012	1,587
BSI (MSA)	0.078	0.183	0.159	0.014	245
BSI (non-MSA)	0.059	0.211	0.164	0.012	1,342
Log Model					
BSI	0.038	0.146	0.113	0.009	1,587
BSI (MSA)	0.053	0.126	0.110	0.010	245
BSI (non-MSA)	0.038	0.146	0.114	0.008	1,342

This table reports summary statistics for markets' baseline search intensity (BSI). In Panel A, BSI is defined using the analytic weights first stage estimates. In Panel B, BSI is defined using the GLS first stage estimates.

Table 1.7: **Market Effect and** $BSI(AW)$

	Linear Model						Log Model					
	VW	VW& IV	GLS	VW	VW& IV	GLS	VW	VW& IV	GLS	VW	VW& IV	GLS
$BSI(AW)$	-0.693***	-0.582**	-0.933***	-0.731***	-0.671***	-0.954***	-0.683***	-0.593***	-0.819***	-0.725***	-0.674***	-0.878***
	(0.176)	(0.248)	(0.071)	(0.186)	(0.241)	(0.070)	(0.144)	(0.194)	(0.069)	(0.152)	(0.190)	(0.065)
HHI	0.048	0.560**	-0.016	0.050	0.513***	-0.003	0.019	0.306**	-0.012	0.021	0.282**	-0.003
	(0.047)	(0.218)	(0.010)	(0.042)	(0.188)	(0.010)	(0.026)	(0.130)	(0.008)	(0.024)	(0.112)	(0.007)
ln(Population)	0.092	0.037	-0.010	0.105**	0.071*	0.000	0.045	0.015	-0.004	0.054*	0.035	0.006
	(0.058)	(0.036)	(0.011)	(0.054)	(0.040)	(0.013)	(0.034)	(0.022)	(0.009)	(0.032)	(0.025)	(0.010)
ln(OwnerUnits)	-0.091	-0.017	0.048***	-0.110**	-0.059*	0.033***	-0.036	0.005	0.034***	-0.049	-0.020	0.022***
	(0.059)	(0.029)	(0.009)	(0.055)	(0.034)	(0.012)	(0.034)	(0.018)	(0.007)	(0.032)	(0.020)	(0.008)
ln(RentUnits)	0.004	0.013	0.003	0.011	0.019*	0.009*	0.002	0.007	-0.000	0.006	0.011	0.004
	(0.011)	(0.012)	(0.005)	(0.010)	(0.011)	(0.006)	(0.007)	(0.007)	(0.004)	(0.007)	(0.007)	(0.004)
ln(MedianRent)	-0.029	0.029	0.034***	0.005	0.048*	0.054***	-0.014	0.017	0.020***	0.005	0.029	0.038***
	(0.023)	(0.025)	(0.009)	(0.023)	(0.028)	(0.010)	(0.015)	(0.017)	(0.006)	(0.016)	(0.018)	(0.007)
ln(MedianValue)	0.075***	0.098***	0.040***	0.060***	0.080***	0.036***	0.047***	0.061***	0.032***	0.038***	0.050***	0.028***
	(0.023)	(0.028)	(0.005)	(0.019)	(0.022)	(0.005)	(0.014)	(0.017)	(0.004)	(0.011)	(0.013)	(0.003)
BachelorsPct	-0.297**	-0.330**	-0.023	-0.271**	-0.303**	-0.001	-0.127	-0.151*	0.005	-0.117	-0.140*	0.004
	(0.151)	(0.146)	(0.039)	(0.133)	(0.129)	(0.038)	(0.087)	(0.087)	(0.032)	(0.078)	(0.078)	(0.028)
WorkerPct	0.056**	0.043**	0.004	0.051**	0.037**	-0.002	0.030**	0.024**	0.006	0.027**	0.020**	0.002
	(0.025)	(0.018)	(0.007)	(0.023)	(0.016)	(0.006)	(0.014)	(0.011)	(0.005)	(0.013)	(0.010)	(0.004)
EmploymentPct	0.029	0.070	-0.048	-0.097	-0.100	-0.089**	-0.016	0.010	-0.055**	-0.092	-0.091	-0.083***
	(0.111)	(0.113)	(0.036)	(0.100)	(0.102)	(0.041)	(0.067)	(0.070)	(0.026)	(0.063)	(0.065)	(0.028)
LaborForcePct	-0.073	-0.144**	-0.054***	-0.036	-0.125**	-0.035	-0.033	-0.073**	-0.036***	-0.013	-0.063*	-0.022
	(0.045)	(0.060)	(0.019)	(0.043)	(0.056)	(0.023)	(0.029)	(0.037)	(0.014)	(0.028)	(0.035)	(0.016)
ln(PercapitaIncome)				-0.028	-0.004	-0.018				-0.015	-0.003	-0.013
				(0.026)	(0.027)	(0.012)				(0.017)	(0.018)	(0.008)
MinorityPct				-0.122***	-0.130***	-0.069***				-0.073***	-0.079***	-0.054***
				(0.028)	(0.027)	(0.009)				(0.017)	(0.017)	(0.006)
Constant	0.216	-0.629	0.183***	0.564**	-0.332	0.288***	-0.536***	-1.025***	-0.615***	-0.338**	-0.846***	-0.550***
	(0.321)	(0.549)	(0.066)	(0.226)	(0.457)	(0.090)	(0.194)	(0.332)	(0.050)	(0.142)	(0.276)	(0.065)
Observations	1586	1586	1586	1586	1586	1586	1586	1586	1586	1586	1586	1586
R^2	0.306	0.040	0.507	0.351	0.135	0.584	0.275	0.062	0.512	0.314	0.140	0.591

Note: VW-Variance Weighted, each market is weighted by the standard error of the market fixed effects, estimated in the first stage.

Huber-White standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.8: **Market Fixed Effects and** $BSI(GLS)$

	Linear Model						Log Model					
	VW	VW& IV	GLS	VW	VW& IV	GLS	VW	VW& IV	GLS	VW	VW& IV	GLS
$BSI(GLS)$	-1.015**	-0.848*	-1.055***	-1.460***	-1.358***	-1.126***	-1.086**	-0.944**	-1.012***	-1.543***	-1.458***	-1.114***
	(0.450)	(0.474)	(0.087)	(0.440)	(0.454)	(0.084)	(0.445)	(0.460)	(0.085)	(0.437)	(0.446)	(0.089)
HHI	-0.035	0.287**	-0.024***	-0.020	0.275**	-0.024***	-0.032	0.147**	-0.018***	-0.021	0.143**	-0.017***
	(0.037)	(0.127)	(0.006)	(0.034)	(0.121)	(0.006)	(0.024)	(0.074)	(0.004)	(0.022)	(0.072)	(0.004)
ln(Population)	0.070*	0.041	-0.015*	0.136***	0.121***	-0.009	0.036	0.021	-0.010*	0.085***	0.078***	-0.004
	(0.040)	(0.031)	(0.009)	(0.048)	(0.045)	(0.009)	(0.025)	(0.021)	(0.006)	(0.032)	(0.030)	(0.006)
ln(OwnerUnits)	-0.044	0.000	0.031***	-0.111***	-0.082**	0.023***	-0.016	0.008	0.021***	-0.065**	-0.050**	0.016***
	(0.037)	(0.025)	(0.006)	(0.041)	(0.033)	(0.007)	(0.022)	(0.017)	(0.004)	(0.026)	(0.023)	(0.005)
ln(RentUnits)	-0.027	-0.025	0.009**	-0.022	-0.019	0.010***	-0.019	-0.018	0.006**	-0.015	-0.014	0.006**
	(0.023)	(0.024)	(0.004)	(0.021)	(0.021)	(0.004)	(0.016)	(0.016)	(0.003)	(0.014)	(0.014)	(0.003)
ln(MedianRent)	-0.085**	-0.047	0.019***	-0.110**	-0.082*	0.025***	-0.060**	-0.039	0.012***	-0.081**	-0.066*	0.016***
	(0.038)	(0.040)	(0.006)	(0.047)	(0.048)	(0.007)	(0.028)	(0.029)	(0.004)	(0.034)	(0.035)	(0.005)
ln(MedianValue)	0.071***	0.080***	0.025***	0.060***	0.067***	0.021***	0.046***	0.051***	0.017***	0.039***	0.043***	0.015***
	(0.018)	(0.019)	(0.003)	(0.015)	(0.016)	(0.003)	(0.012)	(0.012)	(0.002)	(0.010)	(0.010)	(0.002)
BachelorsPct	-0.178	-0.161	-0.041	-0.182	-0.166	-0.046*	-0.098	-0.089	-0.027	-0.101	-0.094	-0.033*
	(0.119)	(0.120)	(0.028)	(0.115)	(0.114)	(0.026)	(0.081)	(0.083)	(0.018)	(0.079)	(0.079)	(0.018)
WorkerPct	0.020	0.015	-0.004	0.003	-0.005	-0.002	0.008	0.005	-0.003	-0.005	-0.009	-0.001
	(0.017)	(0.015)	(0.004)	(0.015)	(0.014)	(0.004)	(0.011)	(0.010)	(0.003)	(0.010)	(0.009)	(0.003)
EmploymentPct	0.048	0.065	0.005	-0.145	-0.158	-0.046*	0.030	0.040	0.008	-0.105	-0.111	-0.028
	(0.112)	(0.115)	(0.024)	(0.141)	(0.142)	(0.026)	(0.081)	(0.083)	(0.016)	(0.101)	(0.102)	(0.018)
LaborForcePct	0.067	0.027	-0.014	-0.034	-0.092*	-0.011	0.057	0.033	-0.007	-0.024	-0.057	-0.004
	(0.051)	(0.056)	(0.013)	(0.044)	(0.051)	(0.016)	(0.035)	(0.038)	(0.009)	(0.031)	(0.036)	(0.011)
ln(PercapitaIncome)				0.092**	0.111**	-0.002				0.072**	0.083***	-0.001
				(0.044)	(0.045)	(0.008)				(0.031)	(0.032)	(0.006)
MinorityPct				-0.078***	-0.083***	-0.029***				-0.048***	-0.051***	-0.020***
				(0.021)	(0.020)	(0.006)				(0.014)	(0.013)	(0.004)
Constant	0.690**	0.186	0.528***	0.271	-0.289	0.606***	-0.167	-0.452**	-0.327***	-0.517***	-0.832***	-0.280***
	(0.284)	(0.363)	(0.049)	(0.268)	(0.373)	(0.068)	(0.189)	(0.227)	(0.033)	(0.180)	(0.238)	(0.048)
Observations	1586	1586	1586	1586	1586	1586	1586	1586	1586	1586	1586	1586
R^2	0.126	0.006	0.621	0.162	0.063	0.640	0.111	0.032	0.627	0.150	0.084	0.639

Note: VW-Variance Weighted, each market is weighted by the standard error of the market fixed effects, estimated in the first stage.

Huber-White standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.9: **Rates and Loan Characteristics**

	Mortgage Rate	
Credit Score (/1000)	-1.458***	(0.003)
Original LTV (%)	0.230***	(0.001)
Debt-to-Income Ratio (%)	0.052***	(0.002)
Original Loan Amount (millions)	-0.674***	(0.002)
Original Term (months/1000)	4.591***	(0.002)
Purchase	-0.104***	(0.000)
Correspondent	0.036***	(0.001)
Retail	0.078***	(0.001)
Single Unit	-0.213***	(0.001)
Single Borrower	-0.000	(0.000)
Ginnie	-0.487***	(0.000)
Constant	4.189***	(0.003)
Observations	3,747,364	
R^2	0.665	

Note: There were 3,862,742 loans in our GSEs and Ginnie Mae dataset, 3,747,364 of which were matched with HMDA. In this regression, we use only loans that are matched.
Huber-White standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.10: **Constructed Price Summary Statistics**

Panel A: Quote Price	Min.	Max.	Mean	SD	Obs.
Quote Price Mean	3.980	4.491	4.230	0.068	1,578
Quote Price 10 pctl	3.656	4.491	4.070	0.091	1,578
Quote Price 20 pctl	3.724	4.491	4.116	0.089	1,578
Quote Price 30 pctl	3.797	4.491	4.152	0.092	1,578
Quote Price 40 pctl	3.797	4.491	4.186	0.091	1,578
Quote Price 50 pctl	3.815	4.631	4.223	0.086	1,578
Quote Price 60 pctl	3.922	4.631	4.258	0.081	1,578
Quote Price 70 pctl	4.003	5.127	4.295	0.079	1,578
Quote Price 80 pctl	4.050	5.127	4.334	0.076	1,578
Quote Price 90 pctl	4.135	6.120	4.385	0.105	1,578
Panel B: Origination Price					
Origination Price Mean	3.961	4.469	4.224	0.068	1,572
Origination Price 10 pctl	3.689	4.384	4.072	0.092	1,572
Origination Price 20 pctl	3.724	4.384	4.116	0.092	1,572
Origination Price 30 pctl	3.724	4.384	4.151	0.095	1,572
Origination Price 40 pctl	3.797	4.456	4.184	0.094	1,572
Origination Price 50 pctl	3.797	4.631	4.220	0.091	1,572
Origination Price 60 pctl	3.922	4.631	4.256	0.085	1,572
Origination Price 70 pctl	3.938	4.700	4.293	0.079	1,572
Origination Price 80 pctl	4.050	4.700	4.330	0.074	1,572
Origination Price 90 pctl	4.050	5.687	4.373	0.088	1,572
Panel C: Lender Average Price					
Lender Avg. Price	4.077	4.606	4.250	0.059	1,578
Lender Avg. Price 10 pctl	3.656	4.491	4.052	0.080	1,578
Lender Avg. Price 20 pctl	3.724	4.491	4.117	0.070	1,578
Lender Avg. Price 30 pctl	3.724	4.491	4.166	0.066	1,578
Lender Avg. Price 40 pctl	3.916	4.491	4.207	0.062	1,578
Lender Avg. Price 50 pctl	4.046	4.606	4.245	0.060	1,578
Lender Avg. Price 60 pctl	4.050	5.127	4.280	0.063	1,578
Lender Avg. Price 70 pctl	4.102	5.127	4.314	0.062	1,578
Lender Avg. Price 80 pctl	4.135	5.127	4.356	0.072	1,578
Lender Avg. Price 90 pctl	4.135	6.120	4.441	0.122	1,578

1. The quote price distribution was constructed based on the number of mortgage applications each lender received. The origination price distribution was constructed based on the number of mortgage loans each lender originated. The lender average price distribution was constructed assuming each lender with at least one mortgage approval in a county gets the same sampling weight.
2. The statistics shown here summarize the distribution of these variables *across* markets.

Table 1.11: **Lender Average Price Mean and** ***BSI***

	Linear Model						Log Model					
	OW	OW	OW& IV	OW	OW	OW& IV	OW	OW	OW& IV	OW	OW	OW& IV
BSI(*AW*)	-0.023	-0.180*	-0.141				-0.009	-0.210*	-0.160			
	(0.096)	(0.094)	(0.096)				(0.122)	(0.119)	(0.122)			
BSI(*GLS*)				-0.120	-0.358**	-0.290*				-0.154	-0.507**	-0.411*
				(0.173)	(0.171)	(0.174)				(0.243)	(0.243)	(0.246)
HHI	-0.023	-0.010	-0.084**	-0.022	-0.010	-0.084**	-0.023	-0.010	-0.084**	-0.022	-0.010	-0.084**
	(0.020)	(0.020)	(0.038)	(0.020)	(0.020)	(0.038)	(0.020)	(0.020)	(0.038)	(0.020)	(0.020)	(0.038)
ln(Population)	0.014	0.070***	0.069***	0.012	0.070***	0.069***	0.015	0.070***	0.069***	0.012	0.070***	0.069***
	(0.013)	(0.016)	(0.015)	(0.013)	(0.016)	(0.015)	(0.013)	(0.016)	(0.015)	(0.013)	(0.016)	(0.015)
ln(OwnerUnits)	-0.028***	-0.074***	-0.075***	-0.026**	-0.075***	-0.075***	-0.028***	-0.075***	-0.075***	-0.026**	-0.075***	-0.075***
	(0.010)	(0.013)	(0.013)	(0.010)	(0.013)	(0.013)	(0.010)	(0.013)	(0.013)	(0.010)	(0.013)	(0.013)
ln(RentUnits)	0.016***	0.011*	0.010*	0.016***	0.010*	0.010*	0.016***	0.011*	0.010*	0.016***	0.010*	0.010*
	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)
ln(MedianRent)	0.133***	0.085***	0.080***	0.133***	0.085***	0.079***	0.133***	0.085***	0.080***	0.133***	0.085***	0.079***
	(0.010)	(0.011)	(0.012)	(0.010)	(0.011)	(0.012)	(0.010)	(0.011)	(0.012)	(0.010)	(0.011)	(0.012)
ln(MedianValue)	-0.017***	-0.017***	-0.020***	-0.017***	-0.017***	-0.020***	-0.017***	-0.017***	-0.020***	-0.017***	-0.018***	-0.020***
	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)
BachelorsPct	-0.438***	-0.449***	-0.431***	-0.443***	-0.448***	-0.431***	-0.435***	-0.446***	-0.429***	-0.442***	-0.447***	-0.431***
	(0.054)	(0.053)	(0.053)	(0.054)	(0.053)	(0.053)	(0.054)	(0.053)	(0.053)	(0.054)	(0.053)	(0.053)
WorkerPct	0.002	-0.009	-0.005	0.001	-0.009	-0.005	0.002	-0.009	-0.005	0.002	-0.009	-0.005
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
EmploymentPct	0.176***	0.070*	0.068	0.182***	0.072*	0.070*	0.175***	0.069*	0.067	0.182***	0.072*	0.070*
	(0.037)	(0.042)	(0.041)	(0.038)	(0.042)	(0.041)	(0.037)	(0.042)	(0.041)	(0.038)	(0.042)	(0.041)
LaborForcePct	-0.118***	-0.197***	-0.188***	-0.116***	-0.196***	-0.188***	-0.119***	-0.197***	-0.189***	-0.117***	-0.197***	-0.188***
	(0.021)	(0.024)	(0.024)	(0.021)	(0.024)	(0.024)	(0.021)	(0.024)	(0.024)	(0.021)	(0.024)	(0.024)
ln(PercapitaIncome)		0.089***	0.084***		0.089***	0.085***		0.089***	0.084***		0.089***	0.085***
		(0.011)	(0.011)		(0.011)	(0.011)		(0.011)	(0.011)		(0.011)	(0.011)
MinorityPct		0.008	0.010		0.005	0.008		0.008	0.010		0.005	0.008
		(0.008)	(0.008)		(0.008)	(0.008)		(0.008)	(0.008)		(0.008)	(0.008)
Constant	3.535***	3.047***	3.160***	3.550***	3.057***	3.170***	3.531***	3.044***	3.157***	3.547***	3.053***	3.167***
	(0.074)	(0.094)	(0.107)	(0.074)	(0.095)	(0.107)	(0.074)	(0.094)	(0.107)	(0.074)	(0.094)	(0.107)
Observations	1578	1578	1578	1578	1578	1578	1578	1578	1578	1578	1578	1578
R^2	0.317	0.348	0.341	0.317	0.348	0.341	0.317	0.348	0.340	0.317	0.348	0.341

Note: OW-Origination Weighted, each market is weighted by the number of originations in that market.

Huber-White standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.12: **Quote Price Mean and *BSI***

	Linear Model						Log Model					
	OW	OW	OW& IV	OW	OW	OW& IV	OW	OW	OW& IV	OW	OW	OW& IV
BSI(AW)	-0.587***	-0.738***	-0.716***				-0.738***	-0.936***	-0.906***			
	(0.133)	(0.131)	(0.133)				(0.170)	(0.167)	(0.170)			
BSI(GLS)				-1.128***	-1.365***	-1.324***				-1.604***	-1.975***	-1.916***
				(0.242)	(0.239)	(0.243)				(0.341)	(0.341)	(0.346)
HHI	-0.132***	-0.120***	-0.162***	-0.132***	-0.120***	-0.166***	-0.132***	-0.120***	-0.162***	-0.133***	-0.120***	-0.166***
	(0.032)	(0.032)	(0.053)	(0.032)	(0.032)	(0.053)	(0.032)	(0.032)	(0.053)	(0.031)	(0.032)	(0.053)
ln(Population)	0.008	0.061***	0.061***	0.006	0.063***	0.063***	0.007	0.062***	0.062***	0.005	0.064***	0.063***
	(0.017)	(0.021)	(0.021)	(0.017)	(0.021)	(0.021)	(0.017)	(0.021)	(0.021)	(0.017)	(0.021)	(0.021)
ln(OwnerUnits)	-0.042***	-0.087***	-0.087***	-0.039***	-0.087***	-0.088***	-0.042***	-0.088***	-0.088***	-0.038***	-0.088***	-0.088***
	(0.013)	(0.017)	(0.017)	(0.013)	(0.017)	(0.017)	(0.013)	(0.017)	(0.017)	(0.014)	(0.017)	(0.017)
ln(RentUnits)	0.020***	0.015*	0.014*	0.019**	0.014*	0.013*	0.020***	0.015*	0.014*	0.019**	0.014*	0.013*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ln(MedianRent)	0.145***	0.096***	0.093***	0.143***	0.096***	0.092***	0.146***	0.096***	0.093***	0.144***	0.096***	0.092***
	(0.013)	(0.015)	(0.015)	(0.013)	(0.015)	(0.015)	(0.013)	(0.015)	(0.015)	(0.013)	(0.015)	(0.015)
ln(MedianValue)	-0.018**	-0.018**	-0.019**	-0.018**	-0.018**	-0.020**	-0.018**	-0.019**	-0.020**	-0.018**	-0.019**	-0.020**
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
BachelorsPct	-0.449***	-0.460***	-0.449***	-0.443***	-0.448***	-0.437***	-0.447***	-0.458***	-0.448***	-0.443***	-0.449***	-0.439***
	(0.074)	(0.074)	(0.075)	(0.074)	(0.074)	(0.075)	(0.074)	(0.074)	(0.075)	(0.074)	(0.074)	(0.075)
WorkerPct	0.017*	0.006	0.009	0.017*	0.007	0.009	0.017*	0.007	0.009	0.017*	0.006	0.009
	(0.010)	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)
EmploymentPct	0.134***	0.034	0.033	0.149***	0.038	0.037	0.137***	0.034	0.033	0.154***	0.040	0.039
	(0.050)	(0.055)	(0.055)	(0.051)	(0.055)	(0.055)	(0.050)	(0.055)	(0.055)	(0.051)	(0.055)	(0.054)
LaborForcePct	-0.147***	-0.226***	-0.221***	-0.144***	-0.224***	-0.219***	-0.147***	-0.227***	-0.222***	-0.144***	-0.225***	-0.220***
	(0.032)	(0.035)	(0.036)	(0.032)	(0.035)	(0.036)	(0.032)	(0.035)	(0.036)	(0.032)	(0.035)	(0.036)
ln(PercapitaIncome)		0.090***	0.087***		0.089***	0.087***		0.091***	0.088***		0.091***	0.088***
		(0.015)	(0.015)		(0.015)	(0.015)		(0.015)	(0.015)		(0.015)	(0.015)
MinorityPct		0.015	0.017*		0.006	0.007		0.013	0.015		0.003	0.005
		(0.010)	(0.010)		(0.010)	(0.010)		(0.010)	(0.010)		(0.010)	(0.010)
Constant	3.802***	3.299***	3.364***	3.824***	3.330***	3.399***	3.801***	3.294***	3.360***	3.818***	3.317***	3.388***
	(0.103)	(0.130)	(0.149)	(0.105)	(0.131)	(0.149)	(0.104)	(0.130)	(0.149)	(0.104)	(0.130)	(0.149)
Observations	1578	1578	1578	1578	1578	1578	1578	1578	1578	1578	1578	1578
R^2	0.271	0.292	0.290	0.274	0.293	0.291	0.271	0.292	0.290	0.274	0.294	0.292

Note: OW-Origination Weighted, each market is weighted by the number of originations in that market.

Huber-White standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.13: **Origination Price Mean and *BSI***

	Linear Model						Log Model					
	OW	OW	OW& IV	OW	OW	OW& IV	OW	OW	OW& IV	OW	OW	OW& IV
BSI(*AW*)	-0.673***	-0.828***	-0.819***				-0.846***	-1.052***	-1.039***			
	(0.139)	(0.137)	(0.139)				(0.176)	(0.174)	(0.177)			
BSI(*GLS*)				-1.254***	-1.517***	-1.498***				-1.782***	-2.197***	-2.169***
				(0.250)	(0.249)	(0.254)				(0.352)	(0.354)	(0.360)
HHI	-0.148***	-0.136***	-0.153***	-0.149***	-0.136***	-0.157***	-0.148***	-0.135***	-0.153***	-0.150***	-0.136***	-0.158***
	(0.033)	(0.033)	(0.056)	(0.033)	(0.033)	(0.056)	(0.033)	(0.033)	(0.056)	(0.033)	(0.033)	(0.056)
ln(Population)	0.010	0.065***	0.065***	0.009	0.067***	0.067***	0.010	0.066***	0.066***	0.008	0.068***	0.068***
	(0.018)	(0.022)	(0.022)	(0.018)	(0.022)	(0.022)	(0.018)	(0.022)	(0.022)	(0.018)	(0.022)	(0.022)
ln(OwnerUnits)	-0.044***	-0.090***	-0.090***	-0.041***	-0.091***	-0.091***	-0.043***	-0.091***	-0.091***	-0.040***	-0.091***	-0.091***
	(0.014)	(0.018)	(0.018)	(0.014)	(0.018)	(0.018)	(0.014)	(0.018)	(0.018)	(0.014)	(0.018)	(0.018)
ln(RentUnits)	0.018**	0.012	0.012	0.016**	0.011	0.011	0.017**	0.012	0.012	0.016**	0.011	0.011
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ln(MedianRent)	0.146***	0.099***	0.097***	0.144***	0.098***	0.097***	0.146***	0.099***	0.097***	0.144***	0.099***	0.097***
	(0.013)	(0.015)	(0.015)	(0.013)	(0.015)	(0.015)	(0.013)	(0.015)	(0.015)	(0.013)	(0.015)	(0.015)
ln(MedianValue)	-0.019**	-0.020**	-0.020**	-0.019**	-0.020**	-0.021**	-0.020**	-0.020**	-0.021**	-0.019**	-0.021**	-0.021**
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)
BachelorsPct	-0.427***	-0.438***	-0.434***	-0.418***	-0.424***	-0.419***	-0.426***	-0.437***	-0.433***	-0.418***	-0.426***	-0.421***
	(0.075)	(0.075)	(0.076)	(0.075)	(0.075)	(0.076)	(0.075)	(0.075)	(0.076)	(0.076)	(0.075)	(0.076)
WorkerPct	0.020**	0.009	0.010	0.020**	0.009	0.011	0.020**	0.009	0.010	0.020**	0.009	0.010
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
EmploymentPct	0.148***	0.043	0.042	0.163***	0.048	0.047	0.152***	0.044	0.043	0.168***	0.050	0.049
	(0.052)	(0.056)	(0.056)	(0.052)	(0.056)	(0.056)	(0.052)	(0.057)	(0.056)	(0.052)	(0.056)	(0.056)
LaborForcePct	-0.143***	-0.221***	-0.219***	-0.141***	-0.219***	-0.217***	-0.143***	-0.222***	-0.220***	-0.140***	-0.220***	-0.218***
	(0.034)	(0.037)	(0.038)	(0.034)	(0.037)	(0.038)	(0.034)	(0.037)	(0.038)	(0.034)	(0.037)	(0.038)
ln(PercapitaIncome)		0.088***	0.087***		0.087***	0.086***		0.090***	0.088***		0.089***	0.088***
		(0.015)	(0.015)		(0.015)	(0.015)		(0.015)	(0.015)		(0.015)	(0.015)
MinorityPct		0.009	0.009		-0.002	-0.001		0.007	0.007		-0.005	-0.004
		(0.010)	(0.010)		(0.010)	(0.010)		(0.010)	(0.010)		(0.010)	(0.010)
Constant	3.819***	3.333***	3.360***	3.838***	3.366***	3.398***	3.818***	3.327***	3.355***	3.831***	3.352***	3.385***
	(0.109)	(0.136)	(0.156)	(0.110)	(0.137)	(0.156)	(0.109)	(0.136)	(0.156)	(0.109)	(0.137)	(0.156)
Observations	1572	1572	1572	1572	1572	1572	1572	1572	1572	1572	1572	1572
R^2	0.266	0.284	0.284	0.268	0.285	0.285	0.266	0.285	0.284	0.269	0.286	0.286

Note: OW-Origination Weighted, each market is weighted by the number of originations in that market.

Huber-White standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.14: **Lender Average Price Percentiles and *BSI***

	10pctl	20pctl	30pctl	40pctl	50pctl	60pctl	70pctl	80pctl	90pctl
Linear Model (Partial Control Set)									
BSI(*AW*)	-0.131	-0.229*	-0.374***	-0.215**	-0.137	-0.030	0.091	0.117	-0.715***
	(0.135)	(0.123)	(0.108)	(0.100)	(0.093)	(0.090)	(0.089)	(0.114)	(0.178)
Linear Model (Complete Control Set)									
BSI(*AW*)	-0.176	-0.341***	-0.485***	-0.344***	-0.317***	-0.267***	-0.175**	-0.157	-0.783***
	(0.138)	(0.125)	(0.111)	(0.100)	(0.093)	(0.085)	(0.083)	(0.108)	(0.176)
BSI(*AW*) IV	-0.165	-0.322**	-0.446***	-0.307***	-0.265***	-0.222**	-0.136	-0.120	-0.766***
	(0.141)	(0.127)	(0.113)	(0.101)	(0.095)	(0.087)	(0.084)	(0.115)	(0.184)
Linear Model (Partial Control Set)									
BSI(*GLS*)	-0.359	-0.481**	-0.686***	-0.398**	-0.219	-0.020	0.184	0.190	-1.351***
	(0.245)	(0.220)	(0.192)	(0.178)	(0.165)	(0.159)	(0.157)	(0.207)	(0.319)
Linear Model (Complete Control Set)									
BSI(*GLS*)	-0.320	-0.630***	-0.881***	-0.615***	-0.539***	-0.461***	-0.299**	-0.293	-1.464***
	(0.253)	(0.227)	(0.202)	(0.182)	(0.169)	(0.156)	(0.151)	(0.199)	(0.317)
BSI(*GLS*) IV	-0.300	-0.596***	-0.811***	-0.549***	-0.448***	-0.383**	-0.230	-0.229	-1.431***
	(0.257)	(0.230)	(0.204)	(0.184)	(0.171)	(0.157)	(0.153)	(0.209)	(0.331)
Log Model (Partial Control Set)									
BSI(*AW*)	-0.166	-0.268*	-0.439***	-0.238*	-0.131	0.006	0.155	0.179	-0.904***
	(0.172)	(0.156)	(0.136)	(0.126)	(0.118)	(0.113)	(0.112)	(0.144)	(0.224)
Log Model (Complete Control Set)									
BSI(*AW*)	-0.211	-0.409**	-0.585***	-0.406***	-0.368***	-0.309***	-0.197*	-0.182	-0.994***
	(0.176)	(0.159)	(0.142)	(0.128)	(0.118)	(0.108)	(0.105)	(0.136)	(0.222)
BSI(*AW*) IV	-0.196	-0.384**	-0.535***	-0.358***	-0.301**	-0.252**	-0.147	-0.135	-0.972***
	(0.179)	(0.162)	(0.144)	(0.129)	(0.120)	(0.110)	(0.107)	(0.145)	(0.233)
Log Model (Partial Control Set)									
BSI(*GLS*)	-0.543	-0.688**	-0.964***	-0.549**	-0.277	0.022	0.328	0.335	-1.879***
	(0.345)	(0.310)	(0.270)	(0.251)	(0.233)	(0.225)	(0.221)	(0.293)	(0.452)
Log Model (Complete Control Set)									
BSI(*GLS*)	-0.477	-0.915***	-1.267***	-0.880***	-0.760***	-0.641***	-0.393*	-0.384	-2.062***
	(0.360)	(0.323)	(0.287)	(0.259)	(0.240)	(0.222)	(0.216)	(0.284)	(0.452)
BSI(*GLS*) IV	-0.449	-0.866***	-1.170***	-0.787***	-0.634***	-0.534**	-0.299	-0.294	-2.016***
	(0.366)	(0.327)	(0.290)	(0.260)	(0.242)	(0.223)	(0.218)	(0.299)	(0.472)

Huber-White standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.15: **Quote Price Percentiles and *BSI***

	10pctl	20pctl	30pctl	40pctl	50pctl	60pctl	70pctl	80pctl	90pctl
Linear Model (Partial Control Set)									
BSI(*AW*)	-0.684***	-0.925***	-1.137***	-1.055***	-1.029***	-0.856***	-0.534***	-0.230*	0.031
	(0.189)	(0.183)	(0.186)	(0.177)	(0.168)	(0.160)	(0.153)	(0.139)	(0.150)
Linear Model (Complete Control Set)									
BSI(*AW*)	-0.702***	-1.001***	-1.237***	-1.153***	-1.150***	-1.020***	-0.773***	-0.484***	-0.285**
	(0.195)	(0.189)	(0.190)	(0.178)	(0.167)	(0.158)	(0.150)	(0.136)	(0.142)
BSI(*AW*) IV	-0.743***	-1.027***	-1.225***	-1.133***	-1.124***	-0.959***	-0.720***	-0.442***	-0.267*
	(0.200)	(0.192)	(0.192)	(0.181)	(0.172)	(0.163)	(0.153)	(0.137)	(0.145)
Linear Model (Partial Control Set)									
BSI(*GLS*)	-1.392***	-1.760***	-2.098***	-1.915***	-1.807***	-1.504***	-0.960***	-0.402*	0.021
	(0.336)	(0.323)	(0.333)	(0.315)	(0.303)	(0.291)	(0.275)	(0.241)	(0.267)
Linear Model (Complete Control Set)									
BSI(*GLS*)	-1.319***	-1.840***	-2.249***	-2.083***	-2.045***	-1.829***	-1.426***	-0.886***	-0.533**
	(0.351)	(0.340)	(0.342)	(0.319)	(0.304)	(0.291)	(0.274)	(0.243)	(0.253)
BSI(*GLS*) IV	-1.387***	-1.881***	-2.223***	-2.044***	-1.996***	-1.721***	-1.333***	-0.810***	-0.500*
	(0.360)	(0.346)	(0.346)	(0.325)	(0.312)	(0.298)	(0.277)	(0.245)	(0.258)
Log Model (Partial Control Set)									
BSI(*AW*)	-0.875***	-1.174***	-1.443***	-1.335***	-1.289***	-1.069***	-0.658***	-0.265	0.077
	(0.241)	(0.231)	(0.236)	(0.224)	(0.213)	(0.204)	(0.193)	(0.175)	(0.190)
Log Model (Complete Control Set)									
BSI(*AW*)	-0.887***	-1.270***	-1.576***	-1.469***	-1.456***	-1.292***	-0.981***	-0.605***	-0.339*
	(0.249)	(0.241)	(0.241)	(0.226)	(0.212)	(0.201)	(0.190)	(0.172)	(0.180)
BSI(*AW*) IV	-0.940***	-1.302***	-1.560***	-1.442***	-1.422***	-1.214***	-0.914***	-0.551***	-0.315*
	(0.256)	(0.244)	(0.244)	(0.230)	(0.219)	(0.207)	(0.193)	(0.174)	(0.183)
Log Model (Partial Control Set)									
BSI(*GLS*)	-2.035***	-2.553***	-3.021***	-2.752***	-2.568***	-2.111***	-1.322***	-0.505	0.118
	(0.474)	(0.456)	(0.470)	(0.446)	(0.429)	(0.411)	(0.387)	(0.338)	(0.376)
Log Model (Complete Control Set)									
BSI(*GLS*)	-1.938***	-2.701***	-3.281***	-3.039***	-2.957***	-2.626***	-2.041***	-1.238***	-0.706**
	(0.499)	(0.483)	(0.486)	(0.454)	(0.432)	(0.415)	(0.389)	(0.344)	(0.360)
BSI(*GLS*) IV	-2.032***	-2.756***	-3.244***	-2.983***	-2.889***	-2.475***	-1.910***	-1.133***	-0.660*
	(0.512)	(0.491)	(0.492)	(0.462)	(0.444)	(0.424)	(0.394)	(0.346)	(0.367)

Huber-White standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.16: **Origination Price Percentiles and *BSI***

	10pctl	20pctl	30pctl	40pctl	50pctl	60pctl	70pctl	80pctl	90pctl
Linear Model (Partial Control Set)									
BSI(*AW*)	-0.734***	-0.914***	-1.180***	-1.194***	-1.087***	-0.886***	-0.573***	-0.252*	0.053
	(0.193)	(0.193)	(0.193)	(0.185)	(0.177)	(0.165)	(0.160)	(0.144)	(0.144)
Linear Model (Complete Control Set)									
BSI(*AW*)	-0.781***	-0.988***	-1.284***	-1.307***	-1.211***	-1.056***	-0.804***	-0.504***	-0.289**
	(0.199)	(0.202)	(0.197)	(0.187)	(0.178)	(0.164)	(0.158)	(0.141)	(0.133)
BSI(*AW*) IV	-0.811***	-1.021***	-1.284***	-1.303***	-1.196***	-1.009***	-0.757***	-0.467***	-0.280**
	(0.204)	(0.205)	(0.200)	(0.190)	(0.183)	(0.169)	(0.160)	(0.143)	(0.135)
Linear Model (Partial Control Set)									
BSI(*GLS*)	-1.467***	-1.742***	-2.135***	-2.160***	-1.901***	-1.577***	-1.035***	-0.459*	0.164
	(0.342)	(0.340)	(0.345)	(0.330)	(0.319)	(0.300)	(0.286)	(0.252)	(0.248)
Linear Model (Complete Control Set)									
BSI(*GLS*)	-1.463***	-1.831***	-2.319***	-2.378***	-2.167***	-1.912***	-1.480***	-0.927***	-0.469**
	(0.357)	(0.361)	(0.354)	(0.335)	(0.321)	(0.301)	(0.284)	(0.252)	(0.234)
BSI(*GLS*) IV	-1.511***	-1.883***	-2.314***	-2.366***	-2.138***	-1.828***	-1.395***	-0.863***	-0.452*
	(0.366)	(0.367)	(0.360)	(0.342)	(0.330)	(0.307)	(0.289)	(0.255)	(0.236)
Log Model (Partial Control Set)									
BSI(*AW*)	-0.936***	-1.165***	-1.493***	-1.513***	-1.363***	-1.105***	-0.709***	-0.298	0.115
	(0.246)	(0.245)	(0.245)	(0.235)	(0.225)	(0.210)	(0.202)	(0.181)	(0.182)
Log Model (Complete Control Set)									
BSI(*AW*)	-0.990***	-1.260***	-1.635***	-1.668***	-1.537***	-1.337***	-1.020***	-0.635***	-0.341**
	(0.254)	(0.257)	(0.250)	(0.237)	(0.226)	(0.209)	(0.199)	(0.178)	(0.169)
BSI(*AW*) IV	-1.028***	-1.302***	-1.634***	-1.662***	-1.517***	-1.277***	-0.960***	-0.588***	-0.328*
	(0.260)	(0.261)	(0.254)	(0.242)	(0.232)	(0.214)	(0.203)	(0.181)	(0.171)
Log Model (Partial Control Set)									
BSI(*GLS*)	-2.135***	-2.535***	-3.065***	-3.101***	-2.703***	-2.218***	-1.434***	-0.600*	0.332
	(0.483)	(0.479)	(0.487)	(0.467)	(0.450)	(0.423)	(0.402)	(0.353)	(0.347)
Log Model (Complete Control Set)									
BSI(*GLS*)	-2.144***	-2.697***	-3.380***	-3.470***	-3.140***	-2.749***	-2.122***	-1.311***	-0.615*
	(0.508)	(0.514)	(0.503)	(0.476)	(0.456)	(0.428)	(0.404)	(0.357)	(0.332)
BSI(*GLS*) IV	-2.209***	-2.768***	-3.370***	-3.451***	-3.098***	-2.631***	-2.004***	-1.221***	-0.592*
	(0.521)	(0.522)	(0.511)	(0.487)	(0.469)	(0.437)	(0.409)	(0.361)	(0.335)

Huber-White standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.17: **Lender's Entry Decision**

	Linear Model				Log Model			
$BSI(AW)$	43.975	(30.591)			56.385	(38.338)		
$BSI(AW) \cdot rate$	-11.251	(7.821)			-14.365	(9.802)		
$BSI(AW) \cdot max(0, r - r_{10})$	24.775^{*}	(14.531)			28.275	(18.224)		
$BSI(AW) \cdot max(0, r - r_{25})$	-36.540^{***}	(13.539)			-42.146^{**}	(16.995)		
$BSI(AW) \cdot max(0, r - r_{50})$	7.315	(10.322)			9.114	(12.982)		
$BSI(AW) \cdot max(0, r - r_{75})$	15.981	(10.092)			19.379	(12.715)		
$BSI(AW) \cdot max(0, r - r_{90})$	4.417	(7.005)			5.442	(8.824)		
$BSI(GLS)$			67.750	(52.264)			97.061	(73.673)
$BSI(GLS) \cdot rate$			-16.907	(13.359)			-24.108	(18.831)
$BSI(GLS) \cdot max(0, r - r_{10})$			26.853	(24.617)			33.680	(34.682)
$BSI(GLS) \cdot max(0, r - r_{25})$			-47.491^{**}	(22.771)			-61.609^{*}	(32.064)
$BSI(GLS) \cdot max(0, r - r_{50})$			17.814	(17.316)			25.105	(24.385)
$BSI(GLS) \cdot max(0, r - r_{75})$			17.193	(16.912)			23.583	(23.811)
$BSI(GLS) \cdot max(0, r - r_{90})$			8.348	(11.732)			11.114	(16.504)
$rate$	4.172^{**}	(2.114)	3.931^{*}	(2.212)	4.166^{**}	(2.072)	3.896^{*}	(2.159)
$max(0, r - r_{10})$	-5.045	(3.924)	-2.799	(4.071)	-4.325	(3.848)	-2.215	(3.971)
$max(0, r - r_{25})$	8.213^{**}	(3.651)	6.196^{*}	(3.759)	7.243^{**}	(3.582)	5.396	(3.664)
$max(0, r - r_{50})$	-1.136	(2.774)	-2.094	(2.849)	-1.083	(2.728)	-2.021	(2.777)
$max(0, r - r_{75})$	-10.268^{***}	(2.703)	-8.809^{***}	(2.773)	-10.046^{***}	(2.662)	-8.666^{***}	(2.702)
$max(0, r - r_{90})$	2.941	(1.876)	2.760	(1.922)	2.987	(1.846)	2.868	(1.871)
ln(Population)	0.017	(0.039)	0.020	(0.039)	0.018	(0.039)	0.020	(0.039)
ln(PercapitaIncome)	0.137^{***}	(0.039)	0.133^{***}	(0.039)	0.136^{***}	(0.039)	0.133^{***}	(0.039)
ln(OwnerUnits)	0.278^{***}	(0.032)	0.277^{***}	(0.032)	0.278^{***}	(0.032)	0.277^{***}	(0.032)
ln(RentUnits)	0.060^{***}	(0.015)	0.059^{***}	(0.015)	0.060^{***}	(0.015)	0.058^{***}	(0.015)
ln(MedianRent)	0.263^{***}	(0.029)	0.269^{***}	(0.029)	0.262^{***}	(0.029)	0.270^{***}	(0.029)
ln(MedianValue)	0.395^{***}	(0.019)	0.404^{***}	(0.019)	0.393^{***}	(0.019)	0.405^{***}	(0.019)
BachelorsPct	-0.883^{***}	(0.130)	-0.854^{***}	(0.130)	-0.879^{***}	(0.130)	-0.848^{***}	(0.130)
MinorityPct	-0.092^{***}	(0.030)	-0.099^{***}	(0.030)	-0.097^{***}	(0.030)	-0.100^{***}	(0.030)
WorkerPct	-0.084^{***}	(0.016)	-0.079^{***}	(0.016)	-0.084^{***}	(0.016)	-0.079^{***}	(0.016)
EmploymentPct	-0.360^{***}	(0.117)	-0.392^{***}	(0.117)	-0.362^{***}	(0.117)	-0.396^{***}	(0.117)
LaborForcePct	-0.275^{***}	(0.062)	-0.274^{***}	(0.062)	-0.274^{***}	(0.062)	-0.276^{***}	(0.062)
Constant	-25.882^{***}	(8.275)	-25.275^{***}	(8.659)	-25.874^{***}	(8.109)	-25.193^{***}	(8.452)
Observations	351063		351063		351063		351063	

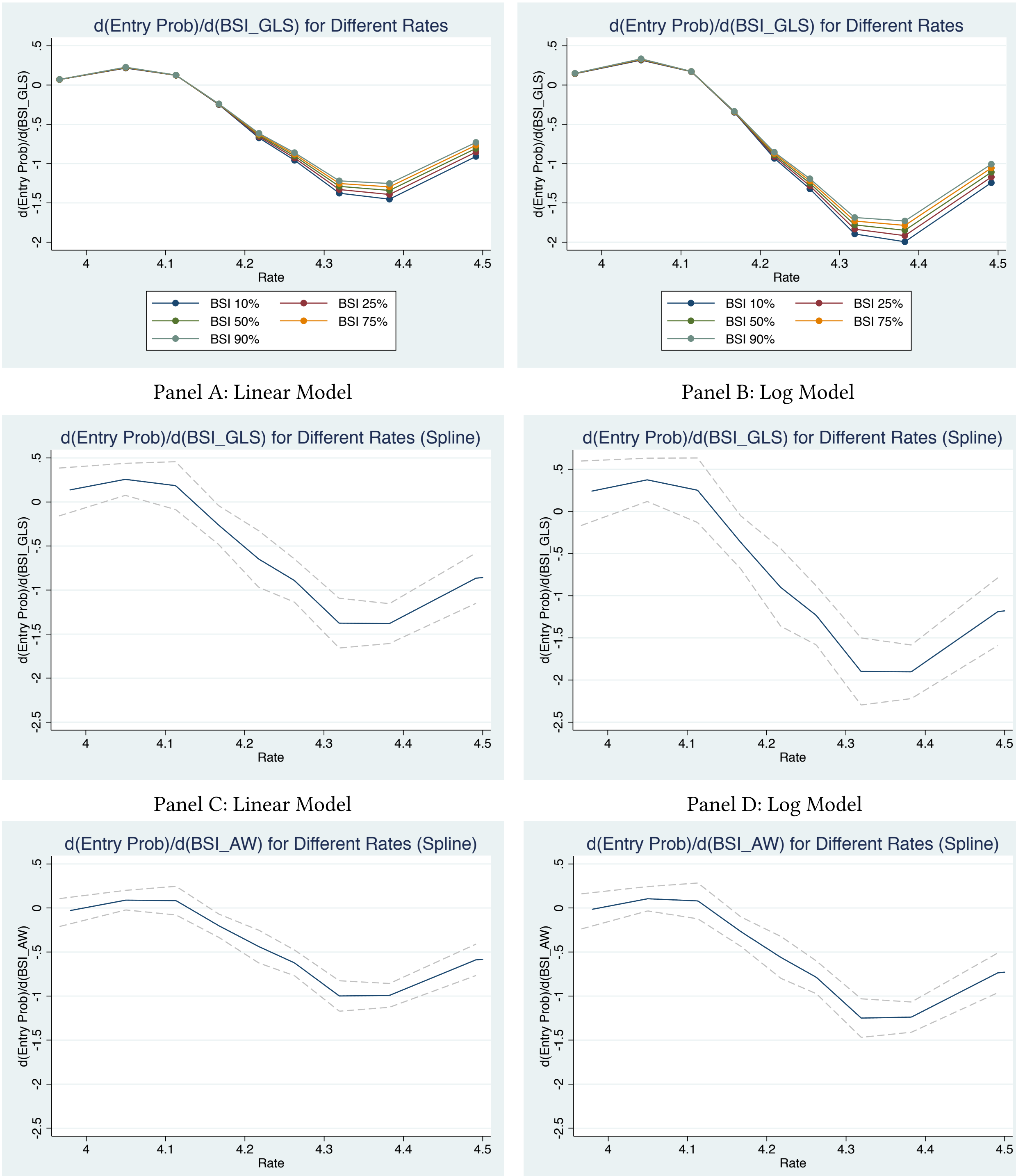
Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.18: **Lender’s Predicted Number of Applications/Originations**

	Linear Model								Log Model							
	Applications		Originations		Applications		Originations		Applications		Originations		Applications		Originations	
$BSI(AW)$	462.997***	(53.557)	569.210***	(60.854)					568.543***	(66.509)	696.920***	(75.226)				
$BSI(AW) \cdot rate$	-112.467***	(13.574)	-139.719***	(15.422)					-137.727***	(16.856)	-170.684***	(19.064)				
$BSI(AW) \cdot max(0, r - r_{10})$	38.735**	(17.527)	91.047***	(19.758)					37.782*	(21.756)	102.457***	(24.445)				
$BSI(AW) \cdot max(0, r - r_{25})$	-16.790	(10.856)	-51.425***	(12.172)					-12.699	(13.480)	-56.297***	(15.105)				
$BSI(AW) \cdot max(0, r - r_{50})$	85.219***	(8.970)	107.331***	(10.323)					103.124***	(11.185)	130.599***	(12.867)				
$BSI(AW) \cdot max(0, r - r_{75})$	25.271**	(11.313)	-13.116	(13.205)					43.012***	(14.190)	-3.576	(16.559)				
$BSI(AW) \cdot max(0, r - r_{90})$	-10.937	(10.534)	41.381***	(14.415)					-24.905*	(13.151)	39.440**	(17.960)				
$BSI(GLS)$					698.501***	(87.346)	844.417***	(98.715)					1024.540***	(125.836)	1224.196***	(141.436)
$BSI(GLS) \cdot rate$					-167.247***	(22.153)	-204.796***	(25.035)					-245.524***	(31.906)	-296.915***	(35.860)
$BSI(GLS) \cdot max(0, r - r_{10})$					-3.492	(29.493)	76.520**	(33.102)					-4.135	(41.904)	106.943**	(46.872)
$BSI(GLS) \cdot max(0, r - r_{25})$					12.058	(19.021)	-43.903**	(21.315)					28.778	(26.424)	-49.315*	(29.600)
$BSI(GLS) \cdot max(0, r - r_{50})$					157.505***	(15.509)	186.860***	(17.812)					215.632***	(21.586)	254.939***	(24.791)
$BSI(GLS) \cdot max(0, r - r_{75})$					48.925**	(19.571)	-5.621	(22.932)					84.806***	(27.380)	12.456	(32.094)
$BSI(GLS) \cdot max(0, r - r_{90})$					-40.362**	(18.087)	47.416*	(24.497)					-72.214***	(25.226)	49.740	(34.210)
$rate$	32.144***	(3.750)	39.625***	(4.273)	29.511***	(3.749)	35.816***	(4.249)	30.886***	(3.645)	37.955***	(4.134)	30.030***	(3.745)	36.007***	(4.220)
$max(0, r - r_{10})$	-10.010**	(4.797)	-24.159***	(5.423)	0.980	(4.943)	-12.234**	(5.562)	-7.581	(4.658)	-21.237***	(5.248)	0.799	(4.870)	-11.892**	(5.460)
$max(0, r - r_{25})$	-1.463	(2.909)	6.855**	(3.264)	-8.034**	(3.125)	0.178	(3.505)	-3.312	(2.822)	4.845	(3.165)	-9.333***	(3.005)	-1.440	(3.369)
$max(0, r - r_{50})$	-17.298***	(2.387)	-21.328***	(2.750)	-20.307***	(2.528)	-23.238***	(2.907)	-16.042***	(2.325)	-19.899***	(2.677)	-18.936***	(2.434)	-21.516***	(2.799)
$max(0, r - r_{75})$	-14.295***	(3.006)	-5.582	(3.507)	-15.520***	(3.185)	-8.120**	(3.731)	-16.514***	(2.945)	-8.324**	(3.436)	-17.113***	(3.083)	-10.441***	(3.613)
$max(0, r - r_{90})$	7.322***	(2.810)	-6.703*	(3.850)	10.979***	(2.950)	-3.445	(4.004)	9.587***	(2.741)	-3.901	(3.748)	12.539***	(2.847)	-1.337	(3.870)
ln(Population)	0.506***	(0.041)	0.479***	(0.047)	0.513***	(0.041)	0.486***	(0.047)	0.509***	(0.041)	0.482***	(0.047)	0.514***	(0.041)	0.487***	(0.047)
ln(PercapitaIncome)	0.319***	(0.041)	0.301***	(0.047)	0.329***	(0.041)	0.313***	(0.047)	0.321***	(0.041)	0.303***	(0.047)	0.332***	(0.041)	0.316***	(0.047)
ln(OwnerUnits)	0.095***	(0.034)	0.103***	(0.039)	0.087**	(0.034)	0.095**	(0.039)	0.093***	(0.034)	0.101**	(0.039)	0.085**	(0.034)	0.093**	(0.039)
ln(RentUnits)	0.025	(0.016)	0.035*	(0.018)	0.026	(0.016)	0.036**	(0.018)	0.025	(0.016)	0.035*	(0.018)	0.026*	(0.016)	0.036**	(0.018)
ln(MedianRent)	0.094***	(0.031)	0.052	(0.035)	0.103***	(0.031)	0.062*	(0.035)	0.096***	(0.031)	0.054	(0.035)	0.104***	(0.031)	0.063*	(0.035)
ln(MedianValue)	0.141***	(0.021)	0.101***	(0.024)	0.147***	(0.021)	0.108***	(0.024)	0.146***	(0.021)	0.107***	(0.024)	0.149***	(0.021)	0.110***	(0.024)
BachelorsPct	0.266**	(0.128)	0.354**	(0.145)	0.293**	(0.128)	0.381***	(0.145)	0.263**	(0.128)	0.351**	(0.145)	0.291**	(0.128)	0.378***	(0.145)
MinorityPct	-0.227***	(0.029)	-0.222***	(0.033)	-0.202***	(0.030)	-0.187***	(0.034)	-0.222***	(0.029)	-0.215***	(0.033)	-0.195***	(0.030)	-0.178***	(0.034)
WorkerPct	0.102***	(0.016)	0.149***	(0.018)	0.106***	(0.016)	0.154***	(0.018)	0.103***	(0.016)	0.150***	(0.018)	0.107***	(0.016)	0.155***	(0.018)
EmploymentPct	0.772***	(0.111)	0.899***	(0.127)	0.742***	(0.111)	0.861***	(0.127)	0.767***	(0.111)	0.891***	(0.127)	0.736***	(0.111)	0.853***	(0.127)
LaborForcePct	0.562***	(0.066)	0.771***	(0.076)	0.538***	(0.066)	0.745***	(0.076)	0.557***	(0.066)	0.766***	(0.076)	0.534***	(0.066)	0.742***	(0.076)
Constant	-143.085***	(14.800)	-171.582***	(16.866)	-134.098***	(14.787)	-157.955***	(16.759)	-138.305***	(14.385)	-165.198***	(16.318)	-136.193***	(14.774)	-158.772***	(16.652)
Observations	351063		351063		351063		351063		351063		351063		351063		351063	

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$



Panel A: Linear Model

Panel B: Log Model

Panel C: Linear Model

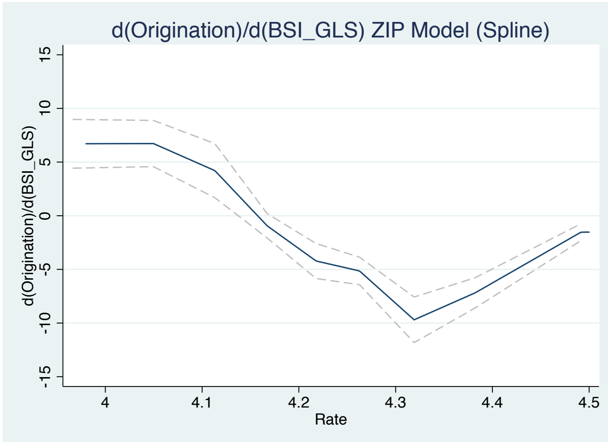
Panel D: Log Model

Panel E: Linear Model

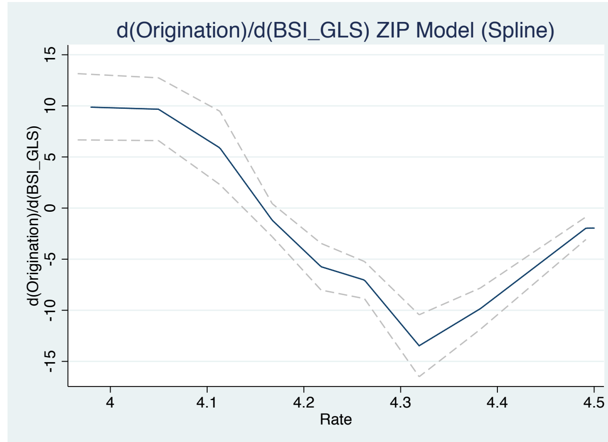
Panel F: Log Model

Figure 1.1: Derivative of Predicted Entry Probability w.r.t. $BSI(GLS)$ or $BSI(AW)$

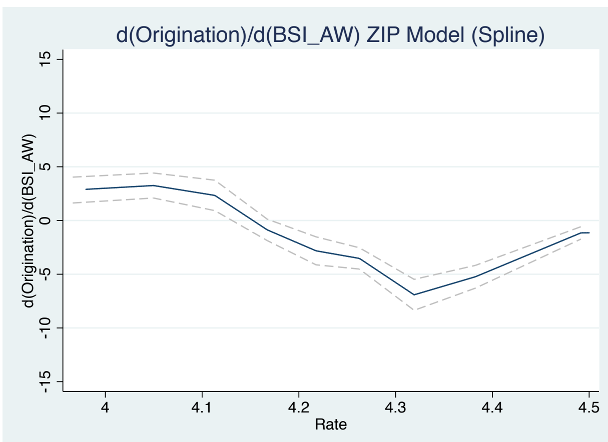
Panels A and B present the derivative of the entry probability with respect to BSI as a function of a lender's lending rate for the linear and log model for different BSI levels. Panels C to F present the spline regression prediction and bootstrapped 95% confidence intervals using $BSI(GLS)$ and $BSI(AW)$ for the linear and log model respectively.



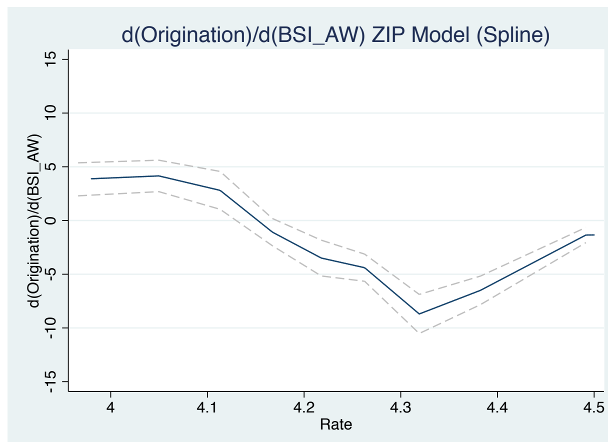
Panel A: Linear Model



Panel B: Log Model



Panel C: Linear Model



Panel D: Log Model

Figure 1.2: Derivative of Predicted Originations w.r.t. $BSI(GLS)$ or $BSI(AW)$

These panels depict BSI's marginal effect on the number of lender's originations as a function of the lender's mortgage rate. The 95% confidence intervals are estimated through bootstrapping.

Chapter 2

Not Pricing to the Market: Evidence from the U.S. Mortgage Industry

2.1 Introduction

Demand for a firm's product and competition from other firms vary across markets, leading to different optimal prices across locations even when marginal costs are the same. Indeed, setting different prices across distinct geographical markets is a very common form of price discrimination. Though most common across national borders, it is also heavily practiced within a country. The U.S. airline industry has historically been very adept to this practice, occasionally even selling a ticket making an extra connection for less if it was originating or terminating in a highly competitive market. Many national and regional retailers also "price to the market" as manufacturers do. ImBev, for example, has regional targets that result in differential pricing across markets. As a testament to the prevalence of geographic price discrimination, anti-trust authorities consider the impact of merger-induced pricing power at the local (rather than national) level, and sometimes require the divestment of retail outlets in specific locations as a condition to permit a merger.[1]

There are, however, constraints to pricing to the market. One constraint is the organizational cost of calculating optimal prices for each location, rather then one national price. This cost can be substantial for some firms, especially those that must set prices for a long list of products. DellaVigna and Gentzkow (2017) consider these costs as the most likely explanation for relative geographic price uniformity among U.S. food, drugstore, and mass merchandise chains. Consumer aversion to overt price discrimination is another constraint. Cavallo, Neiman, and

This chapter is based on joint work with George Deltas. We would like to thank Dan Bernhardt and Sergei Koulayev for very helpful discussions.

[1] The Exxon-Mobil merger is a particularly prominent example of such divestiture, with the companies agreeing to the sale or assignment of 2,431 Exxon and Mobil gas stations in the Northeast and Mid-Atlantic, California, Texas and Guam. When it was agreed, it formed the largest retail divestiture in Federal Trade Commission history ("Exxon/Mobil Agree to Largest FTC Divestiture Ever in Order to Settle FTC Antitrust Charges; Settlement Requires Extensive Restructuring and Prevents Merger of Significant Competing U.S. Assets", FTC, 1999).

Rigobon (2014) suggest that this is the reason for the tendency of prices to be uniform within currency unions and to have become so after an EU member adopts the Euro (Cavallo, Neiman, and Rigobon, 2015, detail the case of Latvia). Price convergence in currency unions can be extreme. For example, Deltas and Desmet (2018) show that geographic price discrimination in the Euro zone for single issues of English language news magazine has dropped to essentially zero after the introduction of the Euro. A feature of these magazines is that they tend to display the single-issue prices for every country on the cover page, making any possible price differences among countries utilizing the same currency readily visible. Consumer aversion to price discrimination may also play a role in non-geographic price uniformity, as illustrated for music (Shiller and Joel Waldfogel, 2011) and for different varieties and flavors of food products (McMillan, 2007).

The U.S. mortgage industry provides an opportunity to assess the importance of organizational costs in limiting "pricing to the market" by firms operating nationally. There is wide variation in mortgage prices: beyond price differences across lenders, there is also a large dispersion in the prices a lender charges to different borrowers. Consumers have no expectations that they will pay the same price. Moreover, it is difficult, if not impossible, to compare the rate that a borrower obtains from a lender to the rate that the lender would charge to an identical borrower in other states. Therefore, consumer push-back cannot be a reason for geographical price uniformity. Though regulatory considerations do impose constraints on price variation within a state, they do not limit price variation across states. Thus, failure to price to the market is not caused by regulation. Costs are also not a likely factor. Though the cost of funds for issuing a mortgage depends on the location and other characteristics of the borrower, there is limited reason that the cost of selling to one type of borrower relative to selling to another type varies from lender to lender. In fact, it is this observation that provides us with the identifying power to test to what extent pricing is influenced by the lenders' operations in multiple markets. This leaves organizational costs as the most likely explanation for departures from pricing to the market, though, as we discuss below, informational factors also seem to play a role.

Mortgage lenders charge different mortgage rates to customers of different characteristics. This differential pricing reflects the credit worthiness of the borrower and the pricing power of the lender with respect to different borrower groups. For each lender, the relationship between mortgage rate and loan/consumer characteristics yields that lender's pricing matrix. An important point for our analysis is the observation that this pricing matrix varies across lenders, as some lenders charge higher rates than others for any given set of characteristics. In fact, one lender may offer lower rates than another lender for some borrowers, while the reverse may be true for other types of borrowers. In addition, a lender may be cheaper than another in one market but more expensive in a different market. In this chapter, we link this heterogeneity in mortgage lenders' pricing to their nation-wide mortgage issuing activity.

A mortgage is a heterogeneous product; its characteristics depend not only on its own attributes, such as the term length and rate adjustability, but also on the attributes of the borrower (e.g., credit score, loan-to-value ratio) and the property (e.g., its location and value). We obtained loan-level data provided by the government sponsored enterprises (GSEs), i.e., Fannie Mae and Freddie Mac, along with Federal Housing Administration (FHA)-insured loan-level data provided by Ginnie Mae, which allows us to control for these attributes. We start by partitioning mortgages into different products based on their most important characteristics, defined as those that enter in the GSEs' pricing matrices. We include in our analysis 13 products that contain enough observations to meaningfully estimate individual lender prices at the state level for the largest lenders. There is still variation in borrower characteristics within those products, which is accounted for in our analysis. But this variation is quite limited, and thus the parametric assumptions in how this heterogeneity is controlled for are not particularly important in determining lender prices.

We then compare the mortgage rates offered for these products by the ten largest multi-state lenders and those offered by "local" lenders operating in a single state. Of direct relevance to whether multi-state lenders price to the market is how the ratio of their rates to those of the local lenders vary with the overall mortgage price level in a state. We find that the rates of the largest multi-state lenders are less responsive to the mortgage price level in a state than the rates of the "local" lenders. In other words, multi-state lenders tend to be partially anchored by their national average price. This effect is particularly pronounced for non-bank lenders. To operationalize these comparisons, we construct a number of measures of a state's price level for each the mortgage products we consider, controlling for residual loan heterogeneity and for lender composition. Our most preferred measure computes a state's price level for the across state variation in prices of lenders that operate in only a very small number of states, but our findings do not hinge on this particular choice.

In the second part of our empirical analysis, we look at the position of lenders in a product's price distribution for each state. We focus on how the lender's position relates to their lending activities in other states, the type of condition it faces in a particular market, and the risk-attributes of the product. This part of our empirical analysis documents three important findings. First, national lenders are less responsive to local market conditions than local lenders. Thus, they are relatively high in the price distribution in low cost states, and relatively low in the price distribution in the high cost states. This echoes our analysis of lenders' average prices described earlier. DellaVigna and Gentzkow (2017) and Hitsch, Hortaçsu, and Lin (2017) report similar results in the U.S. retail industry, where prices and promotions are more homogeneous at the retail chain level than at the market level.

Second, compared to national banks, national non-bank lenders are even less responsive to the local price level. A possible explanation is that banks have more local branches established for mortgage lending and other financial services. This assists them in obtaining soft information from borrowers and evaluating local market conditions. This may also result in increased organizational autonomy in setting interest rates. On the other hand, many non-bank lenders conduct the majority of their business online or through the phone. Typically equipped with better technology than the neighborhood bank branch, these top non-bank lenders provide a more streamlined mortgage origination service. But they are more likely to operate as a unitary organization, limiting systematic differences in pricing across states. Speed and convenience are the two main advantages for non-bank lenders, sometimes in exchange for less flexible, and often higher, prices.

Third, local lenders have lower prices for high risk and refinance loans than lenders active in multiple states. This might also be due to the fact that local lenders have a comparative advantage in collecting and utilizing soft information. In the mortgage lending market, soft information (demographics, future income stability, the quality of the mortgage collateral, etc.) is typically acquired through in-person interactions between the borrower and the loan officer.[2] Several papers have documented the importance of soft information in this market. Keys et al. (2010) point out that the cost of collecting soft information is internalized by lenders to a greater extent when

[2]Our main dataset provided by the GSEs includes hard information (e.g. the borrower's credit score, loan-to-value ratio, and loan purpose, etc.), but does not contain any soft information because investors purchase securitized loans based on hard information only.

a loan is more difficult to securitize, i.e., more likely to stay on the lender's own balance sheet, which eventually leads to lower default rates than similar risk but easier to securitize loans. Relihan (2017) finds that local banks specializing in mortgages appear to use soft information to significantly increase lending to low-quality hard information borrowers, resulting in a lower quality applicant pool for non-local, competing lenders that lack the ability to obtain and process soft information. More broadly, Agarwal et al. (2018) provide evidence on lenders who are geographically close to lottery-winning neighborhoods being more capable of gathering soft information on local shocks and adjusting their lending strategies to mitigate potential bankruptcy risks. Our results indicate that non-local lenders with limited soft information increase their prices to possibly compensate for higher risks. This is true even for the GSE conforming loans.[3] For borrowers that already own a home in the region and looking for refinance, local ties could likely reduce lock risk, i.e., the risk that mortgage rates decline after a previous (higher) rate was locked between lender and borrower, inducing the borrower to find another lender and lock again at a lower rate, thus imposing a net cost on the original lender. Again, non-local lenders unlikely to establish these personal connections raise prices to compensate.

Beyond the literature on spatial price distribution, this chapter is also related to other aspects of the mortgage research. It is relevant to the literature on the importance of soft information in credit markets (Petersen and Rajan 2002; Keys et al. 2009, 2010, 2012; Agarwal and Hauswald 2010; Agarwal et al. 2011; Jiang 2013), to the research on the role different types of mortgage lenders play and their impact on market outcomes (Keys et al. 2009; Andrea and Zazzaro 2011; Rosen 2011), and on the work investigating potential explanations for mortgage price dispersion (Woodward and Hall 2012; Alexandrov and Koulayev 2017).

[3]The vast majority of the GSE securitized loans have a higher than 620 credit score and are regarded as "conforming" loans, as they conform to guidelines established by Fannie Mae and Freddie Mac. These conforming loans are part of the broader category known as conventional loans. Conventional loans are not guaranteed or insured by a government-backed agency such as the Federal Housing Administration (FHA), the Department of Veterans Affairs (VA), the US Department of Agriculture's Rural Development, etc. Note that conventional loans can also be non-conforming. For example, jumbo loans exceed the conforming loan limits and have different underwriting guidelines but are still considered as conventional loans. Hence, within GSE securitized loans, "high risk" is a relative term we use for the products with higher default risk when compared to other products. For FHA-insured loans, we do have borrowers with credit score lower than 620, and we label these loans as "high risk".

2.2 Industry Background

The mortgage industry is a major financial sector in the United States, representing the bulk of household borrowing. The federal government has established several programs to foster mortgage lending and promote home ownership. These programs include the Federal National Mortgage Association (Fannie Mae), the Federal Home Loan Mortgage Corporation (Freddie Mac), and the Government National Mortgage Association (Ginnie Mae). The first two are government-sponsored enterprises (GSEs) while the third is a wholly owned government corporation within the Department of Housing and Urban Development (HUD). The GSEs expand the secondary mortgage market by securitizing mortgages in the form of mortgage-backed securities (MBS). These MBS instruments are eventually sold to institutional investors, such as mutual funds, asset managers, insurance companies, etc. This process facilitates the financing of mortgage lending and allows small lenders with limited capital to reinvest their assets into more lending activities. Similar to the GSEs, Ginnie Mae guarantees the timely payment of principal and interest payments on MBS to the investors. However, while lenders can sell conforming individual loans to the GSEs for cash or swap the loans for the GSEs' MBS, Ginnie Mae does not purchase individual loans from lenders, nor does it issue mortgage-backed securities. Instead, lenders approved by Ginnie Mae originate eligible loans, pool them into securities, and issue Ginnie Mae guaranteed mortgage-backed securities themselves.

The GSEs have a limit on the maximum sized loan they will securitize, known as the "conforming loan limit", which varies by year, location, and property type. In 2014, the limit was 417K in regular areas and 625K in high cost areas for single family loans. Additionally, the borrower's credit score is usually above 620 for conforming loans. Down payment should be at least 20% without private mortgage insurance (PMI) and could be as low as 3% with PMI. Generally, the maximum debt-to-income (DTI) ratio is 43%. However, exceptions can be made for strong compensating factors like high credit scores and low loan-to-value (LTV) ratios. Moreover, to be eligible for a conforming loan, the borrower is required to provide full income and asset documentation.

On the other hand, Ginnie Mae guarantees only securities backed by single-family or multifamily loans insured by government agencies, including the Federal Housing Administration (FHA), the Department of Veterans Affairs (VA), the Department of Agriculture's Rural Development (USDA), and the Department of Housing and Urban Development's Office of Public and Indian Housing (PIH). Moreover, borrower eligibility, loan requirements, and lending terms vary for these government-backed insurers. FHA-insured loans are typically for low-income, low credit score, first-time home buyers with a minimum of 3.5% down payment. An FHA-insured loan also comes with an upfront mortgage insurance premium (MIP) collected when the loan is originated and an annual mortgage insurance premium included in the borrower's monthly mortgage payment. In 2014, the upfront MIP was 1.75% of the base loan amount. The annual MIP was 1.35% for borrowers with a 95.01% or higher LTV ratio and 1.30% for borrowers with a 95.00% or less LTV ratio.[4] VA-insured loans provide a range of benefits for veterans including 0% down payment, no prepayment penalty, limited closing costs, and most importantly, no monthly mortgage insurance, but it has an additional cost in the form of VA funding fees. USDA-insured loans come with rural location restrictions, income limits, and owner-occupation rules. The upside for the borrower that satisfies these requirements is the option of a zero out-of-pocket loan, that is, 0% down payment and financed closing costs.

Unlike the GSEs that specify how the upfront guarantee fee (g-fee) is adjusted according to loan characteristics that reflect loan "riskiness", the upfront g-fee charged by Ginnie Mae is fixed. Since these fees are usually converted by the lender to an ongoing equivalent and reflected in the mortgage rate paid by borrowers, pricing strategies for GSE securitized and Ginnie Mae guaranteed loans are likely different. Another crucial difference between the GSEs and Ginnie Mae is the order in which credit risk is borne in the case of mortgage default. For GSE securitized loans, the mortgage borrower takes the initial credit loss (in the form of house equity), followed by the private mortgage insurance (PMI) company (if any), and then the GSEs. For loans guaranteed by Ginnie Mae, the mortgage borrower is again in the first-loss position, followed by the government entity that insures the loan. However, the lender is expected to bear any credit losses that the government insurer does not cover and Ginnie Mae steps in only when the corporate resources of the lender are exhausted.[5]

Mortgage loans can be provided by different types of financial institutions, the most common of

[4] This annual MIP is conditional on the loan amount being less than $625,500 and loan term longer than 15 years.

[5] The GSEs, Ginnie Mae, and government insurance agencies will not bear the full credit loss if it is proved that the originator or issuer violated the guidelines of their programs.

all are banks that offer a wide range of financial services including deposits, mortgages, business loans, etc. These institutions are highly regulated and usually fund their mortgage originations with deposits or Federal Home Loan Bank advances. They are also more likely to service their own loans and either hold the loans in portfolio or securitize them in pools guaranteed by the GSEs or Ginnie Mae. However, non-bank lenders that specialize in mortgage services thrived post-crisis, especially in the FHA and VA insured loan market. In addition, it is documented that non-banks are more likely to originate mortgages to minority, lower-income, and lower credit-score borrowers (Kim et al. 2018). Another difference between these two types of lending institutions is that while large, national banks tend to have many local branches allowing in-person interactions between the borrower and loan officer, top non-bank lenders are more likely to operate online or through the phone. Therefore, if soft information plays a critical role in mortgage lending, for these major lending institutions at least, banks would be in an advantageous position.

2.3 Data

There are two main datasets we use in this analysis. The first is 30-year fixed-rate, single-family, full-documentation mortgage loans securitized by Fannie Mae (FNMA) and Freddie Mac (FHLMC) in 2014. The share of loans sold to these two GSEs amount to more than 50% of the first-lien loans originated in 2014.[6] The second is 30-year fixed-rate, single-family, full-documentation mortgage loans insured by the Federal Housing Administration (FHA) and guaranteed by Ginnie Mae (GNMA) in 2014.

For each origination securitized by the GSEs, we observe the loan rate, borrower's credit score, the ratio of the loan to the house value (loan-to-value/LTV ratio), the ratio of debt to the borrower's income (debt-to-income/DTI ratio), whether this is a purchase or refinance loan, the origination month, the property's location at the state level, the securitizer (FNMA or FHLMC), loan amount, the number of units, the number of borrowers, third party origination flag, occupation status, property type, and the name of the entity acting in its capacity as a seller of mortgages

[6]In 2014, the GSE share of first-lien originations was 52%, the share of bank portfolio originations was around 27%, FHA/VA originations accounted for another 21%, and the private label origination share was less than 1%.

to the GSEs.[7][8] For the Ginnie Mae guaranteed loans, we observe all the above features except occupation status and property type. However, banks typically offer to each borrower the opportunity to trade-off a lower interest rate with a higher upfront payment known in the industry as "discount points" or "positive points". Another option that works in reverse is to receive a rebate (used to defray loan settlement costs) in exchange for a higher interest rate, which is often known as "negative points". The mortgage rates we observe include point-adjustments (if any), and the specific points chosen by each borrower are not available to us. Thus, a low observed interest rate in our transaction data may reflect a high points payment, and vice versa. In what follows, we assume that conditional on borrower-type, the typical choice of points does not vary across institutions. We also assume that the trade-off between points and interest rates is the same across lenders, that is, lenders may vary in the rates they charge, but not in the points' buy-up or buy-down ratio.

In our analysis, we pool data from both GSEs for 30-year fixed rate, single-family mortgage loans with one or two borrowers for an one unit property classified as a condominium, planned unit development, or a single-family home.[9][10] At the loan level, we have around 1.9 million observations, 60% from Fannie Mae and 40% from Freddie Mac.[11] Our entire dataset has 1,726 lenders. The average GSE lender is active in 4 states and originates 287 loans within a state. In total, we have 6,666 distinct lender-state combinations.

In our Ginnie Mae dataset, the majority of loans were insured by the Federal Housing Administration (57%), followed by the Veterans Administration (30%), and then the U.S. Department of Agriculture's Rural Development (12%).[12] However, due to the differences in the borrower pop-

[7] We dropped observations that have one or more than one of the aforementioned key variables missing.

[8] For most cases, this seller is also the lender of the loan. However, if lender A sells the loans it originated to lender B and lender B later sold these loans to the GSEs, the seller name is recorded as lender B in our dataset.

[9] Loans for property types: co-op, leasehold, manufactured home; properties with 2-4 units, and loans with more than two borrowers, summed up to less than 3.5% of our initial dataset. These loans are likely subject to additional adjustments in guarantee fees. To avoid unnecessary complexities, these loans are not included in the analysis.

[10] "Single-family housing" refers to properties with one to four units, in contrast to "multifamily housing", i.e. properties with five or more units. Multifamily housing usually requires commercial (instead of residential) mortgage loans, which represent a substantially smaller share of the U.S. mortgage market than single-family mortgages. All of our loans are for single-family housing. This should not be confused with the "single-family home" property type, which represents a 1-4 unit property with fee simple ownership.

[11] The U.S. mortgage market is known to be highly fragmented with many small lenders. However, if a lender has less or equal to 10 originations in a particular state for the entire year of 2014, we deem this lender to be inactive in that state. These originations were dropped in the analysis.

[12] The U.S. Department of Housing and Urban Development's Office of Public and Indian Housing also insured 0.34% of the loans in our dataset.

ulation served by these three government agencies, we focus on FHA-insured loans only. Moreover, we restrict our sample to 30-year fixed rate, single-family loans with one or two borrowers for an one unit property to ease the comparison with GSE loans. At the loan level, we have nearly 600K FHA-insured loans originated by 220 lenders. The average Ginnie Mae lender is active in 7 states and originates 323 loans within a state.[13] In total, there are 3,074 distinct lender-state combinations. One might notice that the number of Ginnie Mae lenders is only slightly more than a tenth of the number of GSE lenders. As we discussed previously, this is mainly because lenders cannot directly sell individual loans to Ginnie Mae. Therefore, the lenders we observe in Ginnie Mae's dataset are only those capable of pooling a large of amount of loans either originated in-house or purchased from other smaller lenders.

During the loan origination process, especially for conventional loans that are later sold to the GSEs, pricing is largely determined by a set of key features. Given these borrower and loan characteristics, the loan officer utilizes rate sheets or pricing software to quote a rate to the borrower. In other words, this fairly standardized procedure treats loans with similar characteristics as "homogeneous" products. In light of this, we selected eight mortgage products with comparatively large market shares and focus our analysis to this sample. First, these products all have a loan-to-value ratio between 75% and 80%. Since loans with a higher LTV ratio typically require private mortgage insurance to meet the GSEs' underwriting guidelines, more than 30% of the loans are concentrated in this 5% LTV ratio range. Second, credit score is divided into four brackets: [660,700), [700,740), [740,780), and [780, 820). Fannie Mae's Loan-Level Price Adjustment (LLPA) Matrix and Freddie Mac's Credit Fee in Price Matrix have smaller credit score brackets with 20-point increments.[14] However, we bundled adjacent brackets to increase our sample size for each product. Third, we include both purchase and refinance loan products in our sample. Although for a given set of characteristics, the refinance product market is usually smaller than its purchase product counterparts, it is still a distinct and indispensable market that should be represented. Therefore, we have selected four purchase and four refinance products with LTV ratio

[13] Again, if a lender has less or equal to 10 originations in a particular state for the entire year of 2014, we deem this lender to be inactive in that state. These originations were dropped in the analysis.

[14] The Loan-Level Price Adjustment (LLPA) Matrix and Credit Fee in Price Matrix are used by Fannie Mae and Freddie Mac respectively to charge for potential risk factors in the loans delivered to them. These adjustments are made based on features such as credit score, LTV ratio, loan purpose, occupancy type, product type, etc. Although the two GSEs each have their own matrix, the pricing adjustments are almost identical for these two agencies.

between 75% and 80% and credit score in [660,700), [700,740), [740,780), and [780, 820), respectively. After selection, the 8-product sample consists of more than 545K loan-level observations and 1,713 lenders. The average lender is active in 3.85 states, originates 83 selected loan products within a state, and we are left with 6,602 distinct lender-state combinations.[15]

Due the differences between the GSEs and Ginnie Mae in federal agency status, program structure, risk position, g-fee, etc., we define another set of mortgage products for the FHA-insured loans guaranteed by Ginnie Mae and separate GSE and FHA products for most of our empirical analysis. To facilitate our comparison, we focus on five "homogeneous" mortgage products with relatively large market shares. These products all have an LTV ratio between 95.01% to 96.5% because a lower than 95.01% LTV ratio would decrease the annual MIP and a higher than 96.5% LTV violates FHA's minimum down payment rules and hence is rare in the FHA loan pool. Credit score is divided into five brackets that define our five products: [580,620), [620,660), [660,700), [700,740), and [740,780). We intentionally included the "high risk" bracket [580,620), which is ineligible for GSE securitization. We also dropped all refinance loans since the majority of FHA-insured loans are taken out by first-time home buyers looking for a purchase. The last restriction we impose is that all products have an upfront MIP of 1.75% and an annual MIP of 1.35%. Our selected 5-product sample consists of more than 387K loan-level observations and 209 lenders. The average lender is active in 14 states, originates 133 selected loan products within a state, and we have 2,918 distinct lender-state combinations.[16]

In Table 2.1, we provide descriptive statistics for Fannie Mae and Freddie Mac securitized loans, FHA-insured loans, the 8-product GSE loan sample, and the 5-product FHA loan sample. Comparing the discrete characteristics composition of Fannie Mae and Freddie Mac, we notice no material difference. In terms of mean origination rate, Freddie Mac is 0.05% lower, which is not unexpected when credit score, LTV ratio, and DTI ratio all indicate more credit-worthy borrowers for Freddie Mac. Purchase loans are slightly over-represented in our 8-product sample, mostly because of our LTV ratio restriction between 75% and 80%. Home-owners looking for refinance have usually built up some equity and have a lower LTV ratio. On average, credit score is about 10

[15]Note that there are 13 lenders and 64 lender-state combinations that are not included in our sample because they did not originate any of the 8 products we selected in 2014.

[16]Note that there are 11 lenders and 156 lender-state combinations that are not included in our sample because they did not originate any of the 5 products we selected in 2014.

points higher in our sample due to the fact that credit score is restricted to above 660. Apart from these two features, our 8-product sample matches the GSE loans reasonably well. Comparing the features between the FHA-insured loans and our 5-product sample, there are two distinctions. First, the 5-product sample consists of only purchase loans. Second, the average LTV ratio is 2.5% higher in our sample. Both distinctions result from our sample selection rules that allow only purchase loans with an LTV ratio between 95.01% and 96.5%. However, as expected, the characteristics for the GSE-securitized loans and FHA-insured loans are less comparable. The latter has a significantly higher proportion of purchase, one borrower, and correspondent loans. Moreover, the borrower quality measured by credit score, LTV ratio, and DTI ratio is lower for FHA-insured loans and the loan amount is 40K-50K smaller. The mean origination rate is also lower for FHA-insured loans, but only because this rate does not include the upfront MIP (1.75%) and annual MIP (1.35%). After adding these insurance payments in, FHA-insured loans are on average 1.4% more expensive than the GSE-securitized loans.[17]

In our dataset, non-banks originated 50% of all mortgages securitized by the GSEs and 54% of the FHA-insured loans. In terms of lenders, 47% of the 1,726 GSE lenders and 72% of the 220 FHA lenders were non-banks. To identify potential differences in pricing strategies between banks and non-banks, we match our lenders with the Federal Deposit Insurance Corporation (FDIC)'s member banks, which cover the vast majority of banking institutions in the U.S.[18]

To illustrate the large extent of heterogeneity in lenders' mortgage pricing for different products, we selected some major lenders in the GSE mortgage market and plotted their mortgage price distributions for one purchase and one refinance product, defined previously as Product 3 and Product 7. These plotted rates in Figure 2.1 are the mortgage rate residuals after controlling for all observable borrower and loan characteristics for properties in the state of California with credit scores between 740 and 780 and LTV ratios between 75% and 80%. Our first two panels display price distributions for the top two national GSE lenders: Wells Fargo and J.P. Morgan. Both are highly regulated bank lenders, and their difference in pricing for these two products is not clearly distinguishable. In Panel C and Panel D, when compared to the non-bank lender

[17]In this calculation, we assume the loan prepays in 5 years. Hence, the upfront MIP was divided by 5 before adding to the annual mortgage rate.

[18]The FDIC is a U.S. government corporation providing deposit insurance to depositors in U.S. banks. As of May 1, 2017, the FDIC provided deposit insurance at 5,844 institutions. Note that credit unions are not insured by the FDIC but by the National Credit Union Administration (NCUA), which is also a government agency.

PennyMac, we observe Bank of America pricing more aggressively for the purchase product. Yet for the refinance product, their prices are quite comparable. In Panel E and Panel F, we have two leading non-bank lenders: Nationstar Mortgage and LoanDepot.com, which provide almost identical rate distributions for the purchase product. However, Nationstar Mortgage offers much better rates for the refinance product. In Table 2.2, we present the mean and standard deviation for these price residuals. The similarities and differences mentioned above are reflected in these summary statistics as well.

2.4 Empirical Analysis

2.4.1 Multi-state vs. Local Lenders

In this section, we aim to compare mortgage rates offered by multi-state lenders and local lenders. As discussed previously, the U.S. mortgage market is highly fragmented with many small local lenders, especially for the GSE securitized loan market since small lenders can sell individual loans directly to the GSEs. We define local lenders as those that only operate in one state. Following this definition of "local", among the 1,713 GSE lenders we have, 1,136 (66%) are local lenders. However, these local lenders only originate 6% of the loans in our 8-product sample. The remaining majority of GSE securitized loans were originated by lenders that operate in two or more states. For FHA loans, we have 45 local lenders (22%) that originated merely 1.65% of the loans in our 5-product sample. To capture potential differences in pricing for multi-state and local lenders, we estimate the following equation that separates their state-product fixed effects:

$$P_{ijlst} = \alpha_i X_i + \alpha_t D_t + \alpha_{js}^{multi} D_{js} \cdot D_l^{multi} + \alpha_{js}^{local} D_{js} \cdot D_l^{local} + \epsilon_{ijlst} \tag{2.1}$$

where P_{ijlst} is the product j mortgage rate for borrower i, originated in month t by lender l for a property in state s. X_i is a set of borrower and loan characteristics including the number of borrowers, property type, third-party originator, occupation status, DTI ratio, and loan amount.[19][20]

[19]Property type and occupation status were not provided for the FHA insured loans. Hence, they were not included as controls in the FHA loan regressions.

[20]Among the borrower and loan characteristics we include in our analysis, the number of borrowers, third-party originator, DTI ratio, and loan amount are not directly represented in the GSEs' price adjustment matrices. However,

D_t represents a set of origination month dummies that capture market movements, especially interest rate fluctuations across time. D_{js} is an indicator variable that equals 1 for a product j loan originated for a property located in state s. D_l^{multi} is an indicator variable that equals 1 when lender l is a multi-state lender, and D_l^{local} is an indicator variable that equals 1 when lender l is active in only one state. The coefficients of interest α_{js}^{multi} and α_{js}^{local} represent the relative pricing for multi-state and local lenders for product j in state s. The residual mortgage rate is denote as ϵ_{ijlst}. In Figure 2.2, each point $(X, Y) = (\alpha_{js}^{local}, \alpha_{js}^{multi})$ represents the estimated fixed effects in equation 2.1 for a particular product j, state s combination. We produce these plots for only the GSE loan estimation results because the number of FHA local lenders is too small. To be more specific, the number of states without local lenders for FHA Product 1 to 5 are 40, 29, 29, 31, and 33, respectively.[21][22] Panel A contains all 8 products in our GSE sample. Fitted values are depicted by the solid red line, which is almost parallel to the dotted 45-degree line in green, i.e., the pricing of multi-state lenders closely match that of local lenders for every product-state combination.[23] In Panels B-E, we examine this relationship for purchase, refinance, high-risk, and low-risk products.[24] All plots are similar to what we observe in Panel A except for low-risk products, where multi-state lenders do raise their prices in markets with more expensive local lenders, but the

these variables might still affect mortgage rates. Both the number of borrowers and the DTI ratio can affect the origination loan amount since adding an additional borrower can increase loan limits in a joint mortgage and the DTI ratio for a conventional loan usually cannot exceed 43%. As for loan amount, it could impact mortgage rates through the following three channels: First, loan officers prefer larger loans since origination fees are typically paid as a certain percentage of the total loan size and the cost of underwriting is not affected by loan size as much. Therefore, to compensate for the difference in origination fees, smaller loans are sometimes charged higher rates. Second, borrowers with larger loans might have a systematic way of selecting mortgage points. For example, they may be more affluent and financially sophisticated than those with smaller loans and choose to buy more discount points since these payments are often tax-deductible. However, we do not have data on points to verify these conjectures. Lastly, loan size may affect the lenders' profit when selling these loans to the capital market, often in the form of mortgage backed securities (MBS). Because smaller loans have less prepayment risk, all else being equal, they trade at higher prices in the secondary market. Hence, lenders would charge lower rates to attract borrowers in need of a smaller loan. The effect of loan amount on mortgage rates is ambiguous as there are several forces working in opposite directions. However, our estimation results show that higher loan amounts are charged lower rates. The third-party origination flag indicates whether the loan was originated by a broker, a correspondent lender, or retail. A mortgage broker acts as a middleman between the borrowers and multiple potential lenders. A correspondent lender originates and funds home loans in their own name and sells these loans to larger mortgage lenders after closing the loan. Considering the fact that the borrower's decision to choose brokers, correspondents, or a retail loan officer may be correlated with unobservable features that we could not control for explicitly, the third party originator is included in the regression.

[21]In total, there are 398 points in our figure. This is less than 8 products multiplied by 52 states because local lenders are not available in all state-product combinations.

[22]The plots are very similar when we just use simple averages across product-state for multi-state and local lenders.

[23]The estimated slope for the fitted line is 0.929, $R_2 = 0.71$.

[24]High-risk is defined as credit score lower than 740. Low-risk products have credit scores higher than 740.

magnitude of increase is much smaller compared to other product categories.

Based on our previous definition, all lenders that have originations in more than one state are categorized as multi-state lenders. However, the pricing strategies of top national lenders operating in more than 48 states such as Wells Fargo, J.P. Morgan, and Quicken Loans, etc. might be different from the average multi-state lender, which is active in less than 10 states for the GSE loan market and less than 20 states for the FHA loan market. Hence, another specification is to single out the top 10 lenders (which in total originated more than 40% of the loans in the GSE 8-product sample and 54% of the loans in the FHA 5-product sample) and allow them to have their individual fixed effects at the product-state level:[25]

$$P_{ijlst} = \alpha_i X_i + \alpha_t D_t + \sum_{i=1}^{10} \alpha_{js}^{top_i} D_{js} \cdot D_l^{top_i} + \alpha_{js}^{multi'} D_{js} \cdot D_l^{multi'} + \alpha_{js}^{local} D_{js} \cdot D_l^{local} + \epsilon_{ijlst} \tag{2.2}$$

where P_{ijlst}, X_i, D_t, D_{js}, and D_l^{local} are defined above. $D_l^{top_i}$ indicates whether lender l is a top i lender ($i = 1, 2, \ldots, 10$), $\alpha_{js}^{top_i}$ reflects the state-product level adjustments in pricing for the top i lender, $D_l^{multi'}$ indicates multi-state lenders that are not among the top 10, and $\alpha_{js}^{multi'}$ captures the average state-product pricing for these lenders. In Figure 2.3 Panels A-C, we plot the relationship between $\alpha_{js}^{top_i}$ and α_{js}^{local} for the top 3 banks in the GSE loan market: Wells Fargo, J.P. Morgan, U.S. Bank. In Panels D-F, we plot the top 3 non-bank lenders in the GSE loan market: Quicken Loans, Franklin American Mortgage Company and PennyMac. Each point $(X, Y) = (\alpha_{js}^{local}, \alpha_{js}^{top_i})$ represents the estimated fixed effects in equation 2.2 for every product-state this top i lender is active in. While the prices of the two largest lenders, namely Wells Fargo and J.P. Morgan, have a very high degree of correlation with local lenders, this is not the case for U.S. Bank. In Panel C, the slope of the fitted line is 0.70, much flatter than the slopes in Panels A (0.99) and B (0.97). For non-bank lenders, Quicken Loans seems to have a similar pattern as U.S. Bank, whereas Franklin American Mortgage Company and PennyMac resemble the top two bank lenders. However, when we plot the same relationship for the top 3 banks in the FHA loan market (Panels A-C): Wells

[25]The top 10 GSE lenders that are active in no less than 48 states are (ranked by the number of loans they originated in our sample): 1. Wells Fargo, 2. J.P. Morgan, 3. Quicken Loans, 4. U.S. Bank, 5. Franklin American Mortgage Company, 6. The Branch Banking and Trust Company, 7. Flagstar Bank, 8. PennyMac, 9. Suntrust Mortgage, 10. Bank of America. The top 10 FHA lenders that are active in no less than 47 states are (ranked by the number of loans they originated in our sample): 1. Wells Fargo, 2. PennyMac, 3. Freedom Mortgage Corp., 4. Pingora Loan Servicing, 5. U.S. Bank, 6. Pacific Union Financial, 7. Quicken Loans, 8. Flagstar Bank, 9. J.P. Morgan, 10. Plaza Home Mortgage.

Fargo, U.S. Bank, J.P. Morgan and the top 3 non-bank lenders in the FHA loan market (Panels D-F): PennyMac, Freedom Mortgage Company, and Pingora Loan Servicing in Figure 2.4, it is obvious that the fitted lines are flattened out for all six lenders, especially the non-bank lenders. This suggests that, when compared with their GSE counterparts, the top FHA lenders are less likely to adjust their prices according to local market conditions.

To systematically analyze these relationships and to disentangle the effects for bank and non-bank lenders, we estimated the following equation at the product-state level for the top 10 GSE or FHA lenders:

$$\log\left(\frac{\alpha_{js}^{top_i}}{\alpha_{js}^{local}}\right) = \beta_0 + \beta_1\alpha_s + \beta_2 D_l^{bank} \cdot \alpha_s + \epsilon_{js} \tag{2.3}$$

where $\log(\frac{\alpha_{js}^{top_i}}{\alpha_{js}^{local}})$ is the log ratio of the top lender i and the local lender's product-state level fixed effects we obtained from equation 2.2, α_s denotes state fixed effects or proxies for the average price level in state s, and D_l^{bank} equals 1 if lender l is a bank and 0 otherwise.[26] To obtain state fixed effects or state price mean proxies, we estimate the following equation:

$$P_{ijlst} = \alpha_i X_i + \alpha_t D_t + \alpha_j D_j + \alpha_s D_s + \epsilon_{ijlst} \tag{2.4}$$

where P_{ijlst}, X_i, and D_t are defined previously. We also include product dummies D_j to absorb variations in credit score and loan purpose and state dummies D_s to explain spatial differences in mortgage pricing and local market conditions. The state fixed effects α_s are what we aim to collect as input for equation 2.3. However, since these state fixed effects are very much influenced by the top lenders in the market, to prevent endogenous controls confounding our estimation results, we use several methods to purge the effects of the top lenders $\alpha_{js}^{top_i}$ and local lenders α_{js}^{local} on α_s. We start with including fixed effects for the top 10 lenders (individually) and local lenders when estimating equation 2.4. This produces a second set of state price mean proxies: α_s^2. We construct a third set, α_s^3, by excluding loans originated by the top 10 and local lenders when estimating equation 2.4. For the fourth and fifth set of proxies, α_s^4 and α_s^5, we use only loans originated by lenders that are active in two to five states and two to ten states in our equation 2.4 estimation,

[26]Among the GSE top 10 lenders, only Quicken Loans, Franklin American Mortgage Company, and PennyMac Corp. are non-bank lenders. Among the FHA top 10 lenders, Wells Fargo, U.S. Bank, Flagstar Bank, and J.P. Morgan are banks.

respectively. For ease of notation, we denote the original set of state fixed effects as α_s^1.

Table 2.4 presents the estimation results for equation 2.3 using different state price mean proxies: α_s^k ($k = 1, 2, ..., 5$), in columns (1)-(5). Panels A and B are for GSE lenders only, with the former containing only the top 10 lenders and the latter extending to top 16 lenders. Panels C and D focus on the top 10 and top 16 FHA lenders, respectively. Panel E displays estimation results for equation 2.3 using top 10 GSE lenders and top 10 FHA lenders stacked with standard errors clustered at the lender level. Panel F extends the results to include top 16 GSE and FHA lenders stacked.[27] We are interested in how the state average price level α_s affects the top 10 (or 16) lender and local lender fixed effects ratio $\log(\frac{\alpha_{js}^{top_i}}{\alpha_{js}^{local}})$. For banks, this effect is captured by $(\beta_1 + \beta_2)$. For non-bank lenders, the coefficient of interest is β_1. In nearly all specifications, β_1 is negative and significant, which means the fixed effects ratio decreases as local prices increase. That is, top non-bank lenders in both the GSE and the FHA loan market are more likely to price towards their average market conditions. On the other hand, for top GSE bank lenders, the sum $(\beta_1+\beta_2)$ is positive and significant in columns (1)-(3) for Panels A and B. This implies that top bank lenders actually increase their prices more than local lenders when market prices raise. However, in the last two columns where the state mean price proxy is constructed using only lenders that are active in 2-5 and 2-10 states, the F-test suggests the sum being insignificant. For top FHA bank lenders the sum $(\beta_1+\beta_2)$ is often insignificant, but always negative when significant, contracting the findings for GSE bank lenders. When we stack the GSE and FHA lenders, the sum is negative and significant in columns (1) and (4), implying the fact that top bank lenders also price towards the mean although the magnitude of reversion is smaller than that of non-bank lenders. In other columns, the sum is insignificant, indicating no definite direction of change for the fixed effect ratios.

[27]The 11-16 top GSE lenders added are (ranked by the number of loans they originated in our sample): 11. Green Tree Servicing, 12. CitiMortgage, 13. Nationstar Mortgage, 14. PHH Mortgage Corporation, 15. Loandepot.com, 16. USAA Federal Savings Bank. Among these six lenders, CitiMortgage and USAA Federal Savings Bank are bank lenders, the other four are non-bank lenders. The 11-16 top FHA lenders added are (ranked by the number of loans they originated in our sample): 11. Stearns Lending, 12. Caliber Home Loans, 13. The Branch Banking and Trust Co., 14. First Guaranty Mortgage Co., 15.Sun West Mortgage Co., 16. Bank of America. Among these six added FHA lenders, only The Branch Banking and Trust Co. and Bank of America are banks, the rest are non-bank lenders.

2.4.2 Quantile and Price Regressions

Mortgage rates can be interpreted as each lender having a base price, and then making additional price adjustments based on borrower and loan features. To measure the relative expensiveness of lenders, we control for all observable features that might affect mortgage rates by estimating the following equation for GSE securitized and FHA insured loans separately:

$$P_{ijlst} = \alpha_i X_i + \alpha_j D_j + \alpha_t D_t + \alpha_s D_s + \alpha_l D_l + \epsilon_{ijlst} \tag{2.5}$$

where P_{ijlst}, X_i, D_j, D_t, D_s, and ϵ_{ijlst} are all defined above. Each lender's national base price is estimated by the coefficients of lender dummies D_l, which reflects the relative expensiveness of lender l. The results of this first stage regression is presented in Table 2.3 for GSE and FHA loans separately. As expected, other things equal, purchase loans have lower rates than refinance loans since lenders bear more risk for the latter.[28] A higher credit score obviously reduces mortgage rates. Having an additional borrower or bearing less debt relative to the borrower's income improves the chances of obtaining better GSE loan rates. Moreover, condos, investment properties, and second homes have higher rates than single-family owner-occupied units due to the additional charges specified in the GSEs' pricing adjustment matrices. Loans originated through brokers are significantly cheaper while those originated through correspondents are more expensive. In addition, a larger loan amount likely brings down mortgage rates for both GSE and FHA loans.

After collecting the lender fixed effects α_l and residuals ϵ_{ijlst} from equation 2.5, we construct the average mortgage rate originated by lender l for product j in state s. This purges variations that can be explained by the borrower and loan characteristics included in equation 2.5:

$$\overline{P}_{jls} = \alpha_l + \frac{1}{N_{jls}} \sum_{i=1}^{N_{jls}} \epsilon_{ijlst} \tag{2.6}$$

where N_{jls} is the number of product j loans originated by lender l in state s. For each product j loan originated in state s by lender l, we replace its origination note rate with the average price

[28] Compared to borrowers looking for a purchase loan, refinance borrowers are more likely to switch between lenders or draw-back from the transaction. The cost of "lock-jumping" and other inherent volatility issues are priced in, thus increasing the refinance rates.

$\overline{P}_{jls}$. This creates a price distribution for product j in state s and allows us to rank the lenders according to their relative position in this distribution. Furthermore, we calculate each lender l's price quantile in state s for product j, denoted as q_{jls}. Table 2.5 presents an example of the price quantiles for the 18 lenders that in total originated 100 GSE securitized purchase loans with credit score between 700 and 740 and LTV ratio between 75% and 80% (GSE product 2) in the state of Alaska. In Table 2.6, we show the correlation between GSE and FHA lender quantiles by product. Unsurprisingly, lender quantile correlation is higher within GSE products or FHA products, averaging at 0.439 for GSE products and 0.583 for FHA products. The lender quantile correlation across GSE and FHA products has a much smaller mean of 0.228. This further corroborates the fact that even the same lender prices its GSE and FHA loans differently. For example, a lender offering competitive FHA loan rates might not price its GSE products as aggressively, at least according to its relative ranking in the state. Following a similar logic, for GSE products, the lender quantile correlation within purchase products or refinance products is higher than across these two different loan purposes.

We now examine how product j's price mean in state s interacting with lender characteristics such as lender l's number of active states and bank indicators influence the lender's price quantile q_{jls}. That is, we allow the marginal effect of increasing the state average price on q_{jls} to vary by lender characteristics, or the marginal effect of lender l operating in an additional state to differ according to product j's price mean in state s. Similarly, we interact local lenders' market share for product j in state s with lender characteristics to analyze the marginal effect of local lenders' market share on q_{jls}. We again define local lenders as lenders that are only active in one state. Our basic regression model is:

$$\begin{aligned} q_{jsl} &= \alpha_l D_l + \gamma_1 \bar{P}_{sj} \cdot \log\left(\sum_s D_{ls}\right) + \gamma_2 \bar{P}_{sj} \cdot D_l^{bank} \\ &+ \delta_1 Local_{sj} \cdot \log\left(\sum_s D_{ls}\right) + \delta_2 Local_{sj} \cdot D_l^{bank} + \epsilon_{jsl} \end{aligned} \tag{2.7}$$

where q_{jsl} is the price quantile for a product j loan originated by lender l in state s and D_l is a vector of lender dummies. $\bar{P}_{sj}$ denotes product j's price mean in state s, which is substituted in some variations with the state average price without the national top 20 lenders, or the state average price for lenders that only operate in less than five states to better represent the local market conditions. Another proxy for $\bar{P}_{sj}$ is the state-product fixed effects estimated in the following equation:

$$P_{ijlst} = \alpha_i X_i + \alpha_{js} D_{js} + \alpha_t D_t + \alpha_l D_l + \epsilon_{ijlst} \tag{2.8}$$

where P_{ijlst}, X_i, D_t, and D_l are as defined previously. However, instead of having separate state and product fixed effects, we now include a vector of state-product dummies D_{js} that equals 1 if the loan is for product j in state s. Lender l's extent of national coverage is represented by the log of lender l's total number of active states: $\log\left(\sum_s D_{ls}\right)$. An FDIC-insured institution (bank) is indicated by $D_l^{bank} = 1$, and $Local_{sj}$ is the local lender market share, i.e., the number of local lender product j originations in state s divided by the total number of product j originations in state s.

In addition to the specification above, we are also interested in how the difference between product j's market share in state s and product j's average market share in all states lender l is active in affect lender l's price quantile q_{jls}. For instance, suppose product j has a much larger market share in state s' than the average state lender l is active in, this might result in lender l being more expensive in state s' if lender l's pricing is correlated (or even uniform) across states and thus tailored towards a smaller product j market share. To differentiate this effect across products, we interact the weight differences with indicators for high risk and refinance products. More specifically, we define the market share of product j in state s as the number of product j loans in state s, N_{js}, divided by the total number of loans across all products in state s:

$$w_{js} = \frac{N_{js}}{\sum_{j'} N_{j's}} \tag{2.9}$$

Accordingly, product j's average market share in all states lender l is active in is defined as:

$$\bar{w}_{jl} = \frac{\sum_s D_{ls} \cdot w_{js}}{\sum_s D_{ls}} \tag{2.10}$$

where indicator variable D_{ls} equals 1 if lender l has originated at least one loan in state s and 0 otherwise. That is, the mean of w_{js} across all states where $D_{ls} = 1$. Another version of (2.10) is weighted by the number of loans in each state s:

$$\tilde{w}_{jl} = \frac{\sum_s D_{ls} \cdot w_{js} \cdot N_s}{\sum_s D_{ls} \cdot N_s} \tag{2.11}$$

A third version which does not depend on lender l is the national market share of product j. That is, the total number of product j loans divided by the total number of loans across all states:

$$\bar{w}_j = \frac{\sum_s N_{js}}{\sum_s \sum_{j'} N_{j's}} \tag{2.12}$$

The following regression is based on 2.7, but includes the weight differences we've defined above. Moreover, we added interaction terms between local lenders and high risk or refinance products to distinguish potential differences in pricing between local and non-local lenders for these riskier products:

$$\begin{aligned} q_{jsl} &= \alpha_l D_l + \gamma_1 \bar{P}_{sj} \cdot \log\left(\sum_s D_{ls}\right) + \gamma_2 \bar{P}_{sj} \cdot D_l^{bank} \\ &+ \delta_1 Local_{sj} \cdot \log\left(\sum_s D_{ls}\right) + \delta_2 Local_{sj} \cdot D_l^{bank} \\ &+ \beta_\Delta (w_{js} - \bar{w}_{jl}) + \beta_{|\Delta|} |w_{js} - \bar{w}_{jl}| + \beta_\Delta^{hr} D_{hr} \cdot (w_{js} - \bar{w}_{jl}) + \beta_{|\Delta|}^{hr} D_{hr} \cdot |w_{js} - \bar{w}_{jl}| \\ &+ \beta_\Delta^{refi} D_{refi} \cdot (w_{js} - \bar{w}_{jl}) + \beta_{|\Delta|}^{refi} D_{refi} \cdot |w_{js} - \bar{w}_{jl}| \\ &+ \phi^{hr} (D_{local} \cdot D_{hr}) + \phi^{refi} (D_{local} \cdot D_{refi}) + \epsilon_{jsl} \end{aligned} \tag{2.13}$$

where w_{js} is product j's market share in state s defined in (2.9), $\bar{w}_{jl}$ is product j's average market share in all states lender l is active in as defined in (2.10), but could also be replaced by the weighted version $\tilde{w}_{jl}$ as in (2.11), or product j's national market share $\bar{w}_j$ as in (2.12). We included not only the difference $(w_{js} - \bar{w}_{jl})$, but also its absolute value term to allow for more flexibility in function form. In addition, these differences are interacted with high risk and refinance product indicators. For GSE loans, high risk products are those with credit score less than 740, indicated by $D_{hr} = 1$, refinance products are indicated by $D_{refi} = 1$. For FHA loans, high risk products are those with credit score less than 620. Note that we did not select refinance products for FHA loans because the number of observations that met our selection criteria was quite limited. Lastly, we have the interaction terms between local lenders and high risk or refinance products.

Table 2.7 and Table 2.8 report the results for our lender-state-product level quantile regressions for GSE loans and FHA loans. Columns (1)-(4) show estimates for equation 2.7, but for different $\bar{P}_{sj}$ values. Column (1) starts with product j's price mean in state s. Column (2) excludes the top 20 lenders and column (3) uses only lenders that are active in no more than five states when calculating $\bar{P}_{sj}$. Column (4) substitutes $\bar{P}_{sj}$ with the product-state fixed effects α_{js} estimated in equation 2.8. Columns (5)-(7) add in variables from equation 2.13 with weights $\bar{w}_{jl}$ defined in equation 2.10. Column (8) and (9) replace the the unweighted $\bar{w}_{jl}$ with the weighted $\tilde{w}_{jl}$ and product j's national market share $\bar{w}_j$, respectively. The marginal effect of lender l operating in more states on price quantiles is $\gamma_1 \bar{P}_{sj} + \delta_1 Local_{sj}$, which based on our estimates is always negative. That is, the more states lender l is active in, the cheaper it is compared to other non-local lenders in state s. This effect is larger in absolute value when $\bar{P}_{sj}$ is larger, that is, when product j in state s is more expensive, which again supports the idea that lenders price towards the average state they are active in. For GSE lenders, this effect is larger in absolute value when state s has a smaller local lender market share. For FHA lenders, this effect is actually larger in absolute value when state s has a larger local lender market share since δ_1 is negative in Table 2.8. This could be because GSE lenders' main competitors are not the non-local small lenders, but other multi-state ones. Meanwhile, local lenders with their collected soft information are more likely to compete with multi-state lenders for FHA loans which are usually riskier. In addition, if lender l is a bank, its price quantile is increased by $\gamma_2 \bar{P}_{sj} + \delta_2 Local_{sj}$, which for GSE lenders, is again negative and larger in absolute value when the state-product price mean is higher or when the local market share is smaller. The results are similar for FHA lenders, with the exception of local market share's coefficient δ_2 being insignificant. This indicates that banks have cheaper prices than non-bank lenders in general, and the difference is larger in more expensive markets or in terms of GSE loans, when there are fewer local lenders. To better interpret the weight difference coefficients, we present the coefficient and estimate of an one unit increase in the absolute value of market share differences on price quantiles by product in Table 2.9. For all purchase products, that is, GSE Products 1-4 and FHA Products 1-5, the larger the deviation of product j's local market share from lender l's average product j market share, the more expensive lender l is in this market. This concurs with the idea that lenders' prices are reverting towards the mean. However, we have opposite results for the refinance products represented by GSE Products 5-8. In this case, the larger the deviation, the cheaper lender l is is in the market. Furthermore, the coefficients for the interaction terms between local lender dummies and high risk products are negative and significant for both GSE and FHA loans, despite the definition of "high risk" being more rigid for the latter. This suggests that local lenders provide cheaper rates for high risk loans, perhaps because they are more capable of collecting soft information on the borrowers and utilize this comparative advantage when pricing high risk loans. The same results apply to GSE refinance loans, which can also be viewed as riskier for lenders due to the higher probability of lock jumps.

2.5 Robustness Checks

In this section, we discuss the robustness of our empirical results. As a summary, none of the robustness specifications we explored change our conclusions substantially.

Since local lenders only originated 6% of the loans in our GSE 8-product sample and less than 2% of the loans in our FHA 5-product sample, the local lenders' market share in most states is quite small and often equals 0 for the FHA sample. We start by dropping the controls for local lenders' market share in the estimation of equation 2.7 and 2.13. There are 1,136 local lenders and 577 non-local lenders in the GSE 8-product sample. To further confirm that our results are not driven by local lenders, we use only non-local lenders for the same estimation. Table 2.11 presents the results using 577 non-local lenders only. Note that excluding local lenders in the regression naturally eliminates the interaction terms $D_{local} \cdot D_{hr}$ and $D_{local} \cdot D_{refi}$. In Table 2.12, we report estimation results for the 302 GSE lenders that are active in at least 4 states, which is the average number of states GSE lenders are active in. In Table 2.13, we exclude controls with the local lenders' market share. We discover no material changes in our key results for all these specifications. However, one noticeable pattern is that as we limit our estimation sample to larger lenders, γ_2, the coefficient for the interaction term between the average state-product price and bank dummy decreases in magnitude and loses its significance in some cases. This implies that it is mainly small, local GSE bank lenders driving the negative effect of being a bank lender on price quantiles and this no longer holds true for lenders that are active in more than 3 states.

The same analyses were applied to our FHA 5-product sample. The results for FHA lenders are very consistent along the dimensions we tested.

2.6 Concluding Remarks

There is a substantial amount of lender pricing heterogeneity in the U.S. mortgage market, even after controlling for key borrower and loan characteristics. This chapter explains how these variations in lenders' pricing relate with an array of loan, lender, and market characteristics. We find that major non-bank lenders have more rigid prices across the markets they operate in when compared to their bank counterparts. Moreover, local lenders have cheaper prices for high risk and refinance loans than lenders active in multiple states. Both findings highlight the importance of local branches in collecting soft information and the pricing advantages this brings to the lender. We also discover lenders in the U.S. mortgage market reverting to their average market's pricing as market coverage expands, which echos patterns of uniform pricing for major suppliers in other industries.

2.7 Tables and Figures

Table 2.1: **Descriptive Statistics**

	Fannie Mae	Freddie Mac	8-product	FHA-insured	5-product
Loan Purpose					
Purchase	62.51%	63.10%	71.88%	88.37%	100.00%
Refinance	37.49%	36.90%	28.12%	11.63%	0.00%
Number of Borrowers					
1 Borrower	53.08%	49.70%	49.94%	62.80%	62.60%
2 Borrowers	46.92%	50.30%	50.06%	37.20%	37.40%
Property Type					
Condo	10.95%	8.35%	8.90%	—	—
Planned Unit Development	27.35%	27.57%	30.02%	—	—
Single Family Home	61.69%	64.08%	61.09%	—	—
Third Party Origination Flag					
Broker	9.41%	10.31%	9.55%	12.72%	12.88%
Correspondent	32.28%	35.62%	35.66%	51.58%	53.97%
Retail	58.31%	54.06 %	54.80%	35.71%	33.14%
Occupation Status					
Investment	9.18%	7.73%	7.42%	—	—
Second Home	4.59%	3.86%	6.28%	—	—
Owner-occupied	86.24%	88.41%	86.31%	—	—
Mean Origination Rate (%)	4.55	4.50	4.49	4.23	4.22
Origination Rate S.D.					
Unconditional (%)	0.32	0.29	0.29	0.37	0.35
Conditional (%)	0.23	0.21	0.19	0.33	0.31
Mean Credit Score	740.65	747.17	752.79	680.56	678.49
Mean Loan-to-value ratio (%)	79.72	78.83	79.63	93.98	96.48
Mean Debt-to-income ratio (%)	34.71	34.47	33.86	40.56	40.70
Mean Loan Amount (in thousands $)	217.01	223.30	229.96	176.04	173.82
Number of loans	1,143,396	770,967	545,501	592,678	387,398

1. This table provides descriptive statistics for Fannie Mae and Freddie Mac securitized loans in the first two columns. The third column features the 8-product sample we selected. The fourth and fifth columns presents descriptive statistics for FHA-insured loans and our 5-product sample, respectively. For the discrete characteristics including loan purpose, number of borrowers, property type, third party origination flag, and occupation status, the percentage share of loans with that characteristic is displayed.
2. The conditional origination rate standard deviation controls for month, product, and state fixed effects, along with observable loan/borrower characteristics: loan purpose, number of borrowers, property type, third party origination flag, occupation status, DTI ratio, and loan amount.
3. For FHA-insured loans, the mean origination mortgage rate does not include the upfront mortgage insurance premium nor the annual mortgage insurance premium.

Table 2.2: **Top Lender Product Price in California Summary Statistics**

	Wells Fargo	J.P. Morgan	Bank of America	PennyMac	Nationstar	LoanDepot.com
Product 3 Mean (%)	4.768	4.778	4.594	4.821	4.774	4.770
Product 3 Std. (%)	0.176	0.170	0.115	0.177	0.169	0.149
Observations	1,933	757	135	371	171	360
Product 7 Mean (%)	4.827	4.825	4.784	4.762	4.866	4.697
Product 7 Std. (%)	0.173	0.179	0.179	0.177	0.207	0.180
Observations	1,271	601	138	224	237	257

This table presents Product 3 (FICO$\in$ [740,780), LTV$\in$ (75,80], purchase loan) and Product 7 (FICO$\in$ [740,780), LTV$\in$ (75,80], refinance loan) price residual mean and standard deviation for six of the top lenders nationwide. All price residuals shown here are from the state of California. To purge the effects of all observable borrower and loan characteristics, these residuals are obtained by estimating equation 2.5 using only loans originated for properties in California.

Table 2.3: **Origination Rate and Borrower/Loan Characteristics**

	GSE Loan Rate			FHA Loan Rate	
Product Type			**Product Type**		
2: Purchase & FICO ∈ [700, 740)	-0.201***	(0.001)	2: Purchase & FICO ∈ [620, 660)	-0.236***	(0.003)
3: Purchase & FICO ∈ [740, 780)	-0.285***	(0.001)	3: Purchase & FICO ∈ [660, 700)	-0.314***	(0.003)
4: Purchase & FICO ∈ [780, 820)	-0.296***	(0.001)	4: Purchase & FICO ∈ [700, 740)	-0.360***	(0.003)
5: Refinance & FICO ∈ [660, 700)	0.155***	(0.001)	5: Purchase & FICO ∈ [740, 780)	-0.378***	(0.003)
6: Refinance & FICO ∈ [700, 740)	-0.082***	(0.001)			
7: Refinance & FICO ∈ [740, 780)	-0.192***	(0.001)			
8: Refinance & FICO ∈ [780, 820)	-0.204***	(0.001)			
Number of Borrowers			**Number of Borrowers**		
Two Borrowers	-0.004***	(0.001)	Two Borrowers	0.002	(0.001)
Third Party Origination Flag			**Third Party Origination Flag**		
Broker	-0.035***	(0.001)	Broker	-0.082***	(0.002)
Correspondent	0.011***	(0.001)	Correspondent	0.046***	(0.002)
Property Type					
Condo	0.076***	(0.001)			
Planned Unit Development	-0.015***	(0.001)			
Occupation Status					
Investment	0.488***	(0.001)			
Second Home	0.012***	(0.001)			
Debt-to-Income Ratio	0.001***	(0.000)	**Debt-to-Income Ratio**	0.000***	(0.000)
Loan Amount	-0.003***	(0.000)	**Loan Amount**	-0.007***	(0.000)
Constant	4.720***	(0.008)	Constant	4.454***	(0.012)
Other Controls			**Other Controls**		
Month Fixed Effects	Yes		Month Fixed Effects	Yes	
State Fixed Effects	Yes		State Fixed Effects	Yes	
Lender Fixed Effects	Yes		Lender Fixed Effects	Yes	
Observations	545,501			387,398	
R^2	0.630			0.341	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table reports loan-level estimation results for equation 2.5. The dependent variable is the GSE origination note rate for the first set of estimation results and the FHA origination note rate for the second set of estimation results. Controls include product, origination month, state, lender fixed effects, loan amount, DTI ratio, and indicators for number of borrowers, property type (GSE only), third party origination flag, and occupation status (GSE only). The GSE estimation baseline is one-borrower purchase loans with credit score between 660 and 700, LTV ratio between 75% and 80%, for owner-occupied single family homes in Alaska, originated though retail in January, 2014. The FHA estimation baseline is one-borrower purchase loans with credit score between 580 and 620, LTV ratio between 95.01% and 96.5% in Alaska, originated though retail in January, 2014.

Table 2.4: **Top Lenders vs. Local Lenders**

Panel A: GSE top 10 lenders	(1)		(2)		(3)		(4)		(5)	
β_1	-0.125***	(0.020)	-0.133***	(0.021)	-0.124***	(0.019)	-0.033**	(0.015)	-0.076***	(0.015)
β_2	0.174***	(0.023)	0.185***	(0.025)	0.162***	(0.022)	0.024	(0.017)	0.093***	(0.018)
Constant	0.032***	(0.001)	0.032***	(0.001)	0.032***	(0.001)	0.032***	(0.000)	0.032***	(0.001)
Observations	3,654		3,654		3,654		3,654		3,654	
R^2	0.015		0.015		0.014		0.002		0.008	
$H_0 : \beta_1 + \beta_2 = 0$	0.000		0.001		0.004		0.329		0.104	
Panel B: GSE top 16 lenders	(1)		(2)		(3)		(4)		(5)	
β_1	-0.140***	(0.015)	-0.152***	(0.016)	-0.134***	(0.014)	-0.049***	(0.011)	-0.088***	(0.011)
β_2	0.193***	(0.019)	0.207***	(0.020)	0.185***	(0.018)	0.028*	(0.014)	0.103***	(0.015)
Constant	0.033***	(0.000)	0.033***	(0.000)	0.033***	(0.000)	0.034***	(0.000)	0.033***	(0.000)
Observations	5,631		5,631		5,631		5,631		5,631	
R^2	0.019		0.020		0.020		0.005		0.011	
$H_0 : \beta_1 + \beta_2 = 0$	0.000		0.000		0.000		0.024		0.137	
Panel C: FHA top 10 lenders	(1)		(2)		(3)		(4)		(5)	
β_1	-0.365***	(0.038)	-0.277***	(0.039)	-0.103***	(0.037)	-0.030**	(0.014)	-0.085***	(0.018)
β_2	0.222***	(0.061)	0.226***	(0.064)	0.175***	(0.060)	0.021	(0.021)	0.090***	(0.029)
Constant	-0.041***	(0.002)	-0.041***	(0.002)	-0.041***	(0.002)	-0.045***	(0.002)	-0.040***	(0.002)
Observations	884		884		884		802		882	
R^2	0.102		0.055		0.011		0.006		0.024	
$H_0 : \beta_1 + \beta_2 = 0$	0.003		0.308		0.126		0.581		0.828	
Panel D: FHA top 16 lenders	(1)		(2)		(3)		(4)		(5)	
β_1	-0.329***	(0.032)	-0.252***	(0.033)	-0.109***	(0.029)	-0.008	(0.012)	-0.066***	(0.015)
β_2	0.153***	(0.054)	0.155***	(0.056)	0.154***	(0.049)	-0.036*	(0.018)	0.043*	(0.026)
Constant	-0.038***	(0.002)	-0.038***	(0.002)	-0.038***	(0.002)	-0.042***	(0.002)	-0.037***	(0.002)
Observations	1,374		1,374		1,367		1,243		1,367	
R^2	0.080		0.044		0.011		0.007		0.014	
$H_0 : \beta_1 + \beta_2 = 0$	0.000		0.034		0.254		0.004		0.270	
Panel E: Stacked top 10 lenders	(1)		(2)		(3)		(4)		(5)	
β_1	-0.322***	(0.077)	-0.261***	(0.072)	-0.136*	(0.068)	-0.136***	(0.017)	-0.118***	(0.030)
β_2	0.197*	(0.098)	0.165	(0.098)	0.092	(0.095)	0.047*	(0.026)	0.077	(0.050)
Constant	0.015***	(0.004)	0.015***	(0.004)	0.016***	(0.004)	0.019***	(0.004)	0.017***	(0.004)
Observations	4,538		4,538		4,538		4,456		4,536	
R^2	0.054		0.032		0.010		0.060		0.023	
$H_0 : \beta_1 + \beta_2 = 0$	0.056		0.124		0.373		0.001		0.277	
Panel F: Stacked top 16 lenders	(1)		(2)		(3)		(4)		(5)	
β_1	-0.306***	(0.046)	-0.254***	(0.043)	-0.141***	(0.034)	-0.119***	(0.017)	-0.112***	(0.018)
β_2	0.184**	(0.066)	0.162**	(0.068)	0.129*	(0.063)	0.002	(0.027)	0.062	(0.037)
Constant	0.017***	(0.004)	0.017***	(0.004)	0.019***	(0.004)	0.022***	(0.004)	0.019***	(0.004)
Observations	7,005		7,005		6,998		6,874		6,998	
R^2	0.050		0.031		0.012		0.058		0.022	
$H_0 : \beta_1 + \beta_2 = 0$	0.033		0.098		0.786		0.000		0.118	

Standard errors in parentheses

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

This table presents equation 2.3 estimation results for top 10 GSE lenders in Panel A, top 16 GSE lenders in Panel B, top 10 FHA lenders in Panel C, top 16 FHA lenders in Panel D, top 10 GSE and FHA lenders stacked in Panel E, top 16 GSE and FHA lenders stacked in Panel F. The dependent variable is the log ratio of the product-state level fixed effects for the top lenders $\alpha_{js}^{top_i}$ and local lenders α_{js}^{local}. Columns (1)-(5) use different proxies for the state average price α_s. Column (1) takes state fixed effects α_s^1 directly from equation 2.4. Column (2) collects the state fixed effects from the same equation, but with fixed effects for the top 10 lenders and local lenders included. Column (3) uses α_s^3, construct by running equation 2.4 excluding the top 10 and local lenders. Column (4) and (5) provide results using α_s^4 and α_s^5, obtained from estimating equation 2.4 with only loans from lenders that are active in two to five states and two to ten states, respectively. The last row in both panels shows p-values for the F-test with the null hypothesis $H_0 : \beta_1 + \beta_2 = 0$. For Panels E and F, standard errors are clustered at the lender level.

Table 2.5: **Lender Price Quantiles for GSE Product 2 in Alaska**

Lender	Number of Loans	Price Mean	Price Quantile
Mt. McKinley Bank	3	4.486%	0.02
Denali State Bank	4	4.534%	0.055
State Farm Bank	1	4.586%	0.08
Guild Mortgage Company	1	4.607%	0.09
Quicken Loans	3	4.641%	0.11
First Bank	1	4.643%	0.13
PennyMac	1	4.650%	0.14
USAA Federal Savings Bank	1	4.673%	0.15
Residential Mortgage	4	4.692%	0.175
Wells Fargo Bank	35	4.700%	0.37
U.S. Bank	4	4.703%	0.565
Flagstar Bank	2	4.706%	0.595
PHH Mortgage	2	4.721%	0.615
Alaska USA Federal Credit Union	31	4.765%	0.78
First National Bank Alaska	1	4.770%	0.94
Plaza Home Mortgage	1	4.971%	0.95
United Shore Financial Services	3	5.004%	0.97
Caliber Home Loans	2	5.199%	0.995

This table displays the 18 lenders that originated GSE Product 2 loans (i.e., purchase loans with borrower's credit score between 700 and 740 and LTV ratio between 75% and 80%) in Alaska, along with each lender's number of loans N_{jls}, price mean $\overline{P}_{jls}$, and price quantile q_{jls}.

Table 2.6: **GSE and FHA Lender Quantile Correlation by Product**

	GSE Products								**FHA Products**				
	Prod1	Prod2	Prod3	Prod4	Prod5	Prod6	Prod7	Prod8	Prod1	Prod2	Prod3	Prod4	Prod5
GSE Prod1	1												
GSE Prod2	0.519	1											
GSE Prod3	0.485	0.569	1										
GSE Prod4	0.503	0.572	0.592	1									
GSE Prod5	0.358	0.351	0.329	0.322	1								
GSE Prod6	0.388	0.435	0.436	0.449	0.397	1							
GSE Prod7	0.366	0.415	0.425	0.434	0.377	0.529	1						
GSE Prod8	0.367	0.405	0.432	0.436	0.377	0.522	0.499	1					
FHA Prod1	0.257	0.297	0.297	0.268	0.091	0.065	0.058	-0.002	1				
FHA Prod2	0.333	0.334	0.365	0.341	0.183	0.159	0.141	0.117	0.618	1			
FHA Prod3	0.333	0.354	0.361	0.374	0.198	0.189	0.174	0.158	0.506	0.719	1		
FHA Prod4	0.303	0.318	0.313	0.308	0.197	0.198	0.166	0.147	0.444	0.615	0.680	1	
FHA Prod5	0.275	0.273	0.303	0.293	0.162	0.160	0.137	0.124	0.397	0.587	0.657	0.606	1

This table presents the correlation between GSE and FHA lender quantiles by product. The only correlations insignificant at the 95% confidence level are between FHA Product 1 and GSE Product 6, FHA Product 1 and GSE Product 7, and FHA Product 1 and GSE Product 8.

Table 2.7: **GSE Lender-State-Product Quantile Regressions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
γ_1	-0.047***	-0.041***	-0.043***	-0.051***	-0.047***	-0.047***	-0.048***	-0.046***	-0.048***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)
γ_2	-0.088***	-0.098***	-0.083***	-0.066***	-0.088***	-0.088***	-0.036*	-0.039*	-0.034*
	(0.019)	(0.020)	(0.022)	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)
δ_1	0.101***	0.099***	0.098***	0.106***	0.102***	0.102***	0.099***	0.101***	0.098***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
δ_2	0.189***	0.167***	0.187***	0.216***	0.189***	0.189***	0.228***	0.226***	0.235***
	(0.059)	(0.060)	(0.061)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)
β_Δ					0.293***	0.293***	0.345***	0.323***	0.396***
					(0.083)	(0.083)	(0.108)	(0.103)	(0.094)
$\beta_{\lvert\Delta\rvert}$						-0.017	0.428***	0.628***	0.253**
						(0.125)	(0.139)	(0.131)	(0.119)
β_Δ^{hr}							-1.042***	-0.898***	-0.955***
							(0.195)	(0.182)	(0.165)
$\beta_{\lvert\Delta\rvert}^{hr}$							0.811***	0.777***	0.740***
							(0.224)	(0.207)	(0.191)
β_Δ^{refi}							1.053***	0.938***	0.971***
							(0.190)	(0.184)	(0.175)
$\beta_{\lvert\Delta\rvert}^{refi}$							-2.524***	-2.239***	-1.965***
							(0.192)	(0.182)	(0.170)
ϕ_{hr}							-0.037***	-0.036***	-0.044***
							(0.008)	(0.008)	(0.008)
ϕ_{refi}							-0.058***	-0.058***	-0.031***
							(0.007)	(0.007)	(0.008)
Constant	1.111***	1.055***	1.043***	1.130***	1.106***	1.107***	1.038***	1.013***	1.032***
	(0.041)	(0.043)	(0.044)	(0.042)	(0.041)	(0.042)	(0.047)	(0.047)	(0.047)
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,117	38,117	37,861	38,117	38,117	38,117	38,117	38,117	38,117
R^2	0.377	0.376	0.377	0.377	0.377	0.377	0.379	0.382	0.382

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table presents the estimation results for equations 2.7 and 2.13 using the GSE 8-product sample. The dependent variable is the lender-state-product price quantile q_{jsl}. Column (1) uses product j's price mean in state s for $\bar{P}_{sj}$. Column (2) substitutes the state-project price mean $\bar{P}_{sj}$ with the state-project price mean without the national top 20 lenders. Column (3) substitutes $\bar{P}_{sj}$ with the state-project average price for lenders that only operate in less than six states. Note that column (3) has slightly fewer observations because some state-product combinations do not have loans originated from lenders that are active in less than six states. Column (4) substitutes $\bar{P}_{sj}$ with the product-state fixed effects α_{js} estimated in equation 2.8. Column (5) adds the difference in product j's market share in state s and product j's average market share in all states lender l is active in as defined in 2.10. Column (6) adds the aforementioned difference's absolute value. Column (7) distinguishes the effects for high risk and refinance products. Column (8) replaces $\bar{w}_{jl}$ with the weighted $\tilde{w}_{jl}$. Column (9) replaces $\bar{w}_{jl}$ with product j's national market share $\bar{w}_j$.

Table 2.8: **FHA Lender-State-Product Quantile Regressions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
γ_1	-0.115***	-0.111***	-0.048***	-0.126***	-0.115***	-0.114***	-0.116***	-0.118***	-0.118***
	(0.005)	(0.005)	(0.004)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
γ_2	-0.100***	-0.074**	-0.069***	-0.121***	-0.100***	-0.104***	-0.100***	-0.100***	-0.097***
	(0.037)	(0.036)	(0.024)	(0.042)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
δ_1	-0.064***	-0.029*	-0.044***	-0.092***	-0.064***	-0.067***	-0.067***	-0.067***	-0.066***
	(0.016)	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
δ_2	-0.048	0.075	0.122	-0.079	-0.048	-0.037	-0.025	-0.025	-0.027
	(0.125)	(0.127)	(0.128)	(0.125)	(0.125)	(0.124)	(0.125)	(0.125)	(0.125)
β_Δ					0.041	0.027	0.050	0.012	0.056
					(0.104)	(0.104)	(0.105)	(0.104)	(0.099)
$\beta_{\lvert\Delta\rvert}$						0.580***	0.608***	0.595***	0.536***
						(0.155)	(0.155)	(0.153)	(0.148)
β_Δ^{hr}							-1.797**	-1.610**	-1.911**
							(0.886)	(0.783)	(0.792)
$\beta_{\lvert\Delta\rvert}^{hr}$							1.353	1.247	1.250
							(1.078)	(0.994)	(1.000)
ϕ_{hr}							-0.150**	-0.150**	-0.138**
							(0.059)	(0.059)	(0.060)
Constant	2.207***	2.122***	1.264***	2.441***	2.208***	2.184***	2.212***	2.232***	2.228***
	(0.064)	(0.063)	(0.042)	(0.077)	(0.064)	(0.064)	(0.076)	(0.077)	(0.077)
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,265	11,252	10,003	11,265	11,265	11,265	11,265	11,265	11,265
R^2	0.473	0.471	0.478	0.470	0.473	0.473	0.474	0.474	0.474

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table presents the estimation results for equations 2.7 and 2.13 using the FHA 5-product sample. The dependent variable is the lender-state-product price quantile q_{jsl}. Column (1) uses product j's price mean in state s for $\bar{P}_{sj}$. Column (2) substitutes the state-project price mean $\bar{P}_{sj}$ with the state-project price mean without the national top 20 lenders. Column (3) substitutes $\bar{P}_{sj}$ with the state-project average price for lenders that only operate in less than six states. Note that column (3) has slightly fewer observations because some state-product combinations do not have loans originated from lenders that are active in less than six states. Column (4) substitutes $\bar{P}_{sj}$ with the product-state fixed effects α_{js} estimated in equation 2.8. Column (5) adds the difference in product j's market share in state s and product j's average market share in all states lender l is active in as defined in 2.10. Column (6) adds the aforementioned difference's absolute value. Column (7) distinguishes the effects for high risk and refinance products. Column (8) replaces $\bar{w}_{jl}$ with the weighted $\tilde{w}_{jl}$. Column (9) replaces $\bar{w}_{jl}$ with product j's national market share $\bar{w}_j$.

Table 2.9: **Price Quantiles and Market Share Differences**

	$w_{js} - \bar{w}_{jl} > 0$		$w_{js} - \bar{w}_{jl} < 0$	
GSE Product, LTV ∈ (75%, 80%]	Coefficient	Estimates	Coefficient	Estimates
1: Purchase & FICO ∈ [660, 700) 2: Purchase & FICO ∈ [700, 740)	β_Δ+$\beta_{\lvert\Delta\rvert}$+ β^{hr}_Δ+$\beta^{hr}_{\lvert\Delta\rvert}$	0.543	-β_Δ+$\beta_{\lvert\Delta\rvert}$-$\beta^{hr}_\Delta$+$\beta^{hr}_{\lvert\Delta\rvert}$	1.936
3: Purchase & FICO ∈ [740, 780) 4: Purchase & FICO ∈ [780, 820)	β_Δ+$\beta_{\lvert\Delta\rvert}$	0.773	-β_Δ+$\beta_{\lvert\Delta\rvert}$	—
5: Refinance & FICO ∈ [660, 700) 6: Refinance & FICO ∈ [700, 740)	β_Δ+$\beta_{\lvert\Delta\rvert}$+ β^{hr}_Δ+$\beta^{hr}_{\lvert\Delta\rvert}$+ β^{refi}_Δ+$\beta^{refi}_{\lvert\Delta\rvert}$	-0.929	-β_Δ+$\beta_{\lvert\Delta\rvert}$-$\beta^{hr}_\Delta$+$\beta^{hr}_{\lvert\Delta\rvert}$-$\beta^{refi}_\Delta$+$\beta^{refi}_{\lvert\Delta\rvert}$	-1.640
7: Refinance & FICO ∈ [740, 780) 8: Refinance & FICO ∈ [780, 820)	β_Δ+$\beta_{\lvert\Delta\rvert}$+ β^{refi}_Δ+$\beta^{refi}_{\lvert\Delta\rvert}$	-0.698	-β_Δ+$\beta_{\lvert\Delta\rvert}$-$\beta^{refi}_\Delta$+$\beta^{refi}_{\lvert\Delta\rvert}$	-3.493
FHA Product, LTV ∈ (95%, 96.5%]	Coefficient	Estimates	Coefficient	Estimates
1: Purchase & FICO ∈ [580, 620)	β_Δ+$\beta_{\lvert\Delta\rvert}$+ β^{hr}_Δ+$\beta^{hr}_{\lvert\Delta\rvert}$	—	-β_Δ+$\beta_{\lvert\Delta\rvert}$-$\beta^{hr}_\Delta$+$\beta^{hr}_{\lvert\Delta\rvert}$	3.709
2: Purchase & FICO ∈ [620, 660) 3: Purchase & FICO ∈ [660, 700) 4: Purchase & FICO ∈ [700, 740) 5: Purchase & FICO ∈ [740, 780)	β_Δ+$\beta_{\lvert\Delta\rvert}$	0.659	-β_Δ+$\beta_{\lvert\Delta\rvert}$	0.558

This table presents how the market share differences can affect lender l's relative expensiveness for each product. Since we include absolute value terms, the effects for $w_{js} - \bar{w}_{jl} > 0$ and $w_{js} - \bar{w}_{jl} < 0$ are shown separately. For both cases, we show the effect of an one unit increase in the absolute value of market share differences on price quantiles. GSE products estimates are produced with the estimation results in Table 2.7 Column (7). FHA products estimates are produced with the estimation results in Table 2.8 Column (7).

Table 2.10: **GSE Lender-State-Product Quantile Regressions (No Local Share Controls)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
γ_1	-0.054***	-0.050***	-0.052***	-0.056***	-0.054***	-0.054***	-0.055***	-0.053***	-0.055***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
γ_2	-0.096***	-0.109***	-0.097***	-0.072***	-0.096***	-0.096***	-0.047**	-0.049**	-0.045**
	(0.019)	(0.020)	(0.021)	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)
β_Δ					0.263***	0.262***	0.295***	0.312***	0.365***
					(0.083)	(0.083)	(0.108)	(0.103)	(0.094)
$\beta_{\lvert\Delta\rvert}$						-0.192	0.236*	0.492***	0.123
						(0.125)	(0.139)	(0.131)	(0.119)
β_Δ^{hr}							-1.098***	-1.015***	-1.029***
							(0.195)	(0.182)	(0.166)
$\beta_{\lvert\Delta\rvert}^{hr}$							0.814***	0.758***	0.723***
							(0.225)	(0.207)	(0.192)
β_Δ^{refi}							1.157***	1.033***	1.053***
							(0.190)	(0.185)	(0.175)
$\beta_{\lvert\Delta\rvert}^{refi}$							-2.460***	-2.156***	-1.899***
							(0.193)	(0.182)	(0.171)
ϕ_{hr}							-0.037***	-0.037***	-0.044***
							(0.008)	(0.008)	(0.008)
ϕ_{refi}							-0.056***	-0.056***	-0.031***
							(0.007)	(0.007)	(0.008)
Constant	1.220***	1.190***	1.192***	1.215***	1.216***	1.226***	1.159***	1.134***	1.153***
	(0.040)	(0.041)	(0.042)	(0.042)	(0.040)	(0.041)	(0.047)	(0.046)	(0.046)
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,117	38,117	37,861	38,117	38,117	38,117	38,117	38,117	38,117
R^2	0.374	0.373	0.374	0.373	0.374	0.374	0.379	0.379	0.379

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table presents the estimation results for equations 2.7 and 2.13 using the GSE 8-product sample without controls that contain the local market share $Local_{sj}$. See Table 2.7 footnotes for column specifications.

Table 2.11: **GSE Lender-State-Product Quantile Regressions (Lenders Active in >1 State)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
γ_1	-0.053***	-0.044***	-0.044***	-0.058***	-0.052***	-0.052***	-0.048***	-0.046***	-0.047***
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
γ_2	-0.032	-0.065***	-0.065**	0.007	-0.032	-0.032	-0.031	-0.033	-0.028
	(0.023)	(0.024)	(0.025)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
δ_1	0.099***	0.097***	0.097***	0.105***	0.099***	0.099***	0.099***	0.101***	0.099***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
δ_2	0.214***	0.187***	0.198***	0.233***	0.214***	0.214***	0.223***	0.221***	0.228***
	(0.063)	(0.064)	(0.065)	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)
β_Δ					0.292***	0.292***	0.345***	0.323***	0.360***
					(0.085)	(0.085)	(0.111)	(0.106)	(0.104)
$\beta_{\lvert\Delta\rvert}$						-0.014	0.429***	0.628***	0.321**
						(0.128)	(0.143)	(0.135)	(0.130)
β_Δ^{hr}							-1.041***	-0.898***	-1.006***
							(0.200)	(0.186)	(0.185)
$\beta_{\lvert\Delta\rvert}^{hr}$							0.810***	0.776***	0.657***
							(0.230)	(0.212)	(0.209)
β_Δ^{refi}							1.053***	0.938***	1.002***
							(0.195)	(0.189)	(0.187)
$\beta_{\lvert\Delta\rvert}^{refi}$							-2.524***	-2.239***	-2.216***
							(0.197)	(0.186)	(0.184)
Constant	1.219***	1.150***	1.148***	1.249***	1.214***	1.215***	1.162***	1.131***	1.137***
	(0.048)	(0.050)	(0.052)	(0.050)	(0.048)	(0.049)	(0.057)	(0.056)	(0.056)
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,657	31,657	31,401	31,657	31,657	31,657	31,657	31,657	31,657
R^2	0.314	0.313	0.314	0.314	0.314	0.314	0.318	0.319	0.319

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

There are 577 lenders active in more than 1 state (non-local) and 1,136 local lenders in the GSE 8-product sample. This table presents the estimation results for equations 2.7 and 2.13 using the 577 non-local lenders only. See Table 2.7 footnotes for column specifications.

Table 2.12: **GSE Lender-State-Product Quantile Regressions (Lenders Active in >3 States)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
γ_1	-0.055***	-0.046***	-0.046***	-0.062***	-0.055***	-0.055***	-0.053***	-0.050***	-0.052***
	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
γ_2	0.010	-0.037	-0.041	0.058**	0.010	0.010	0.010	0.007	0.011
	(0.026)	(0.027)	(0.028)	(0.026)	(0.026)	(0.026)	(0.025)	(0.025)	(0.025)
δ_1	0.100***	0.100***	0.098***	0.106***	0.101***	0.101***	0.100***	0.102***	0.100***
	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
δ_2	0.187***	0.156**	0.167**	0.202***	0.187***	0.186***	0.191***	0.193***	0.193***
	(0.070)	(0.071)	(0.071)	(0.069)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)
β_{Δ}					0.298***	0.297***	0.343***	0.377***	0.386***
					(0.089)	(0.089)	(0.116)	(0.111)	(0.111)
$\beta_{\|\Delta\|}$						-0.089	0.307**	0.504***	0.206
						(0.132)	(0.149)	(0.143)	(0.139)
β_{Δ}^{hr}							-1.062***	-0.954***	-1.091***
							(0.209)	(0.196)	(0.197)
$\beta_{\|\Delta\|}^{hr}$							0.908***	0.893***	0.828***
							(0.242)	(0.225)	(0.226)
β_{Δ}^{refi}							1.035***	0.920***	0.946***
							(0.203)	(0.198)	(0.198)
$\beta_{\|\Delta\|}^{refi}$							-2.366***	-2.132***	-2.056***
							(0.206)	(0.197)	(0.198)
Constant	1.266***	1.192***	1.198***	1.299***	1.260***	1.266***	1.229***	1.195***	1.214***
	(0.053)	(0.055)	(0.057)	(0.054)	(0.053)	(0.054)	(0.062)	(0.062)	(0.062)
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,172	28,172	27,916	28,172	28,172	28,172	28,172	28,172	28,172
R^2	0.289	0.287	0.288	0.289	0.289	0.289	0.293	0.293	0.293

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

There are 302 lenders active in more than 3 states in the GSE 8-product sample. This table presents the estimation results for equations 2.7 and 2.13 using these 302 lenders only. See Table 2.7 footnotes for column specifications.

Table 2.13: **GSE Lender-State-Product Quantile Regressions**

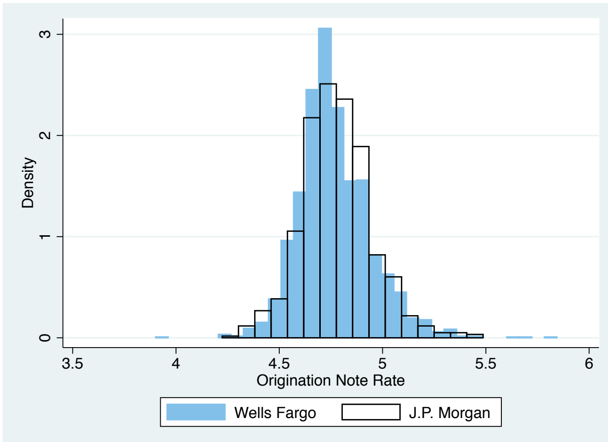
(Lenders Active in >3 States & No Local Market Share Controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
γ_1	-0.062***	-0.055***	-0.056***	-0.066***	-0.062***	-0.063***	-0.060***	-0.058***	-0.059***
	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
γ_2	0.002	-0.047*	-0.054**	0.054**	0.002	0.002	0.002	-0.000	0.003
	(0.025)	(0.026)	(0.027)	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
β_{Δ}					0.269***	0.267***	0.297**	0.369***	0.364***
					(0.089)	(0.089)	(0.117)	(0.111)	(0.111)
$\beta_{\lvert\Delta\rvert}$						-0.269**	0.113	0.341**	0.040
						(0.132)	(0.149)	(0.143)	(0.139)
β_{Δ}^{hr}							-1.118***	-1.071***	-1.180***
							(0.210)	(0.197)	(0.198)
$\beta_{\lvert\Delta\rvert}^{hr}$							0.918***	0.869***	0.808***
							(0.242)	(0.226)	(0.227)
β_{Δ}^{refi}							1.127***	1.000***	1.017***
							(0.203)	(0.199)	(0.198)
$\beta_{\lvert\Delta\rvert}^{refi}$							-2.310***	-2.067***	-1.999***
							(0.206)	(0.197)	(0.198)
Constant	1.397***	1.356***	1.378***	1.401***	1.392***	1.409***	1.373***	1.339***	1.356***
	(0.052)	(0.054)	(0.055)	(0.054)	(0.052)	(0.053)	(0.062)	(0.061)	(0.061)
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,172	28,172	27,916	28,172	28,172	28,172	28,172	28,172	28,172
R^2	0.285	0.284	0.285	0.284	0.285	0.285	0.289	0.289	0.289

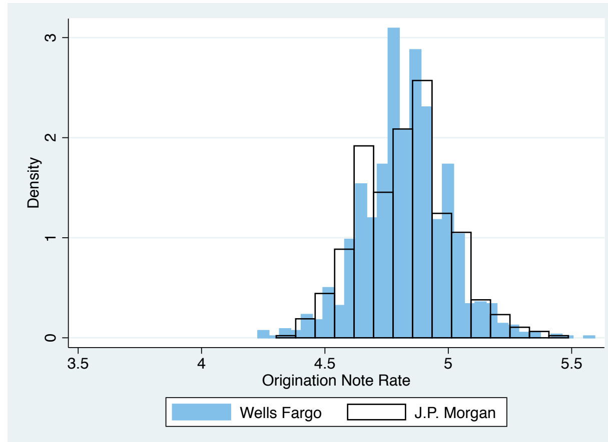
Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

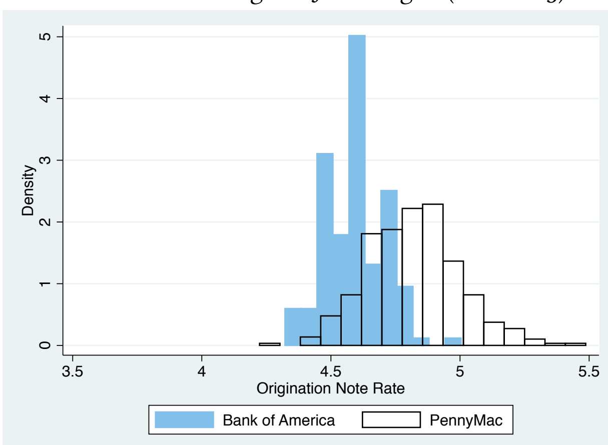
There are 302 lenders active in more than 3 states in the GSE 8-product sample. This table presents the estimation results for equations 2.7 and 2.13 without controls that contain the local market share $Local_{sj}$ using these 302 lenders only. See Table 2.7 footnotes for column specifications.



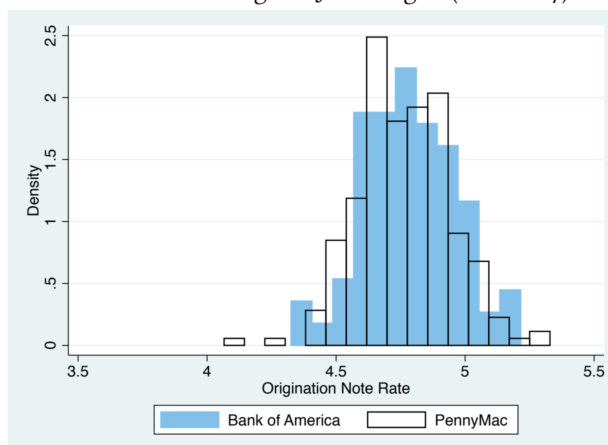
Panel A: Wells Fargo vs J.P. Morgan (Product 3)



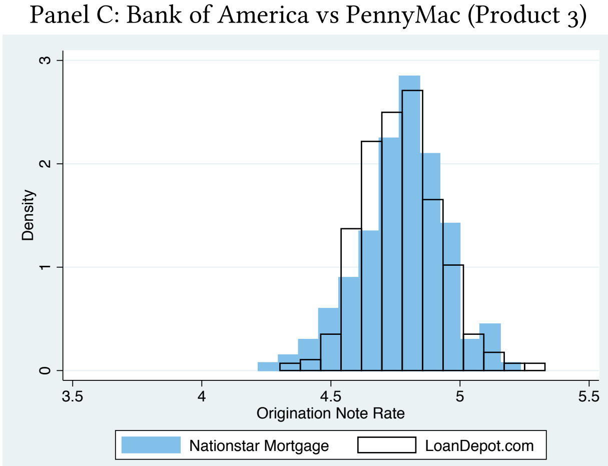
Panel B: Wells Fargo vs J.P. Morgan (Product 7)



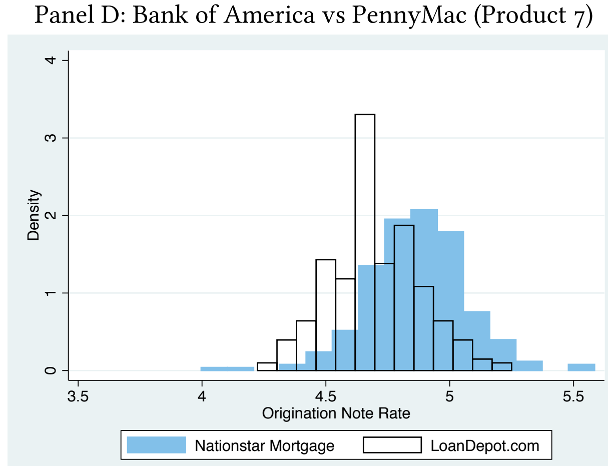
Panel C: Bank of America vs PennyMac (Product 3)



Panel D: Bank of America vs PennyMac (Product 7)



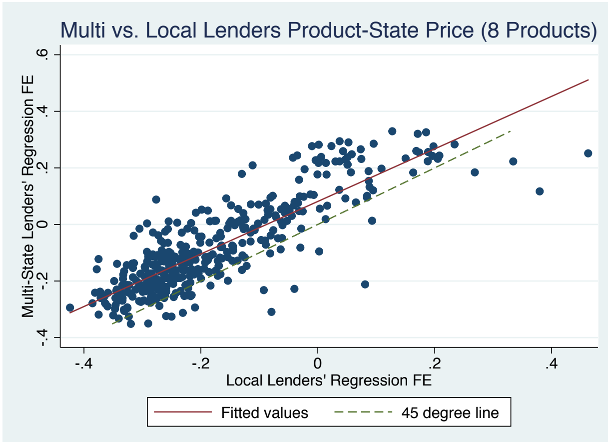
Panel E: Nationstar vs LoanDepot.com (Product 3)



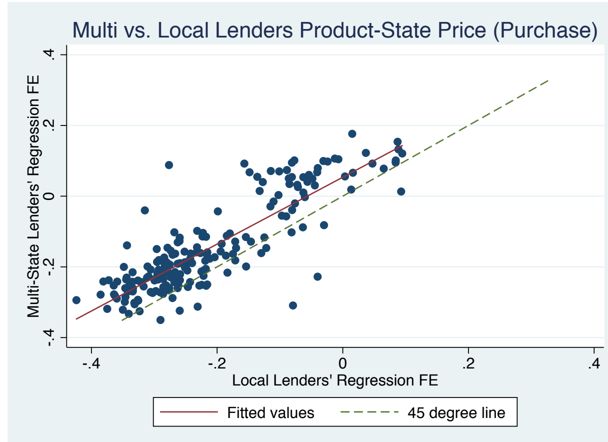
Panel F: Nationstar vs LoanDepot.com (Product 7)

Figure 2.1: Lender Product Price Distribution Comparison

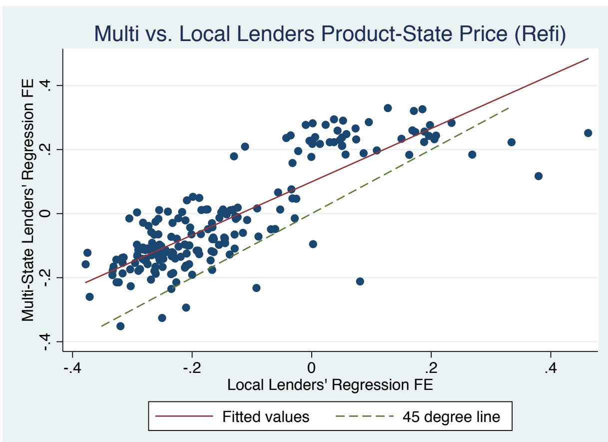
This figure compares the price distributions for Product 3 (FICO ∈ [740,780), LTV ∈ (75,80], purchase loan) and Product 7 (FICO ∈ [740,780), LTV ∈ (75,80], refinance loan) between six major GSE lenders, namely Wells Fargo and J.P. Morgan in panels A and B, Bank of America and PennyMac in panels C and D, and Nationstar Mortgage and LoanDepot.com in panels E and F. These plotted rates are the mortgage rate residuals after controlling for all observable borrower and loan characteristics for properties in the state of California.



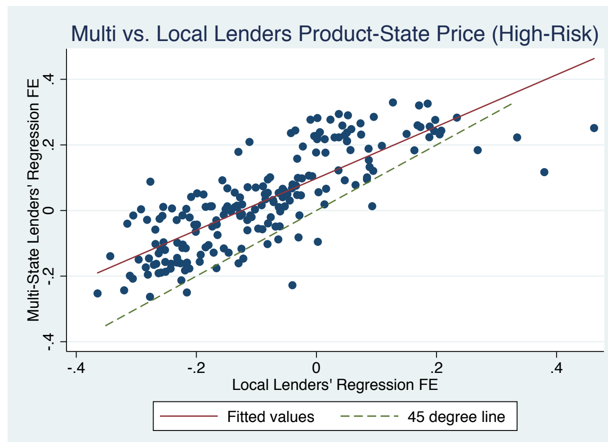
Panel A: All 8 Products



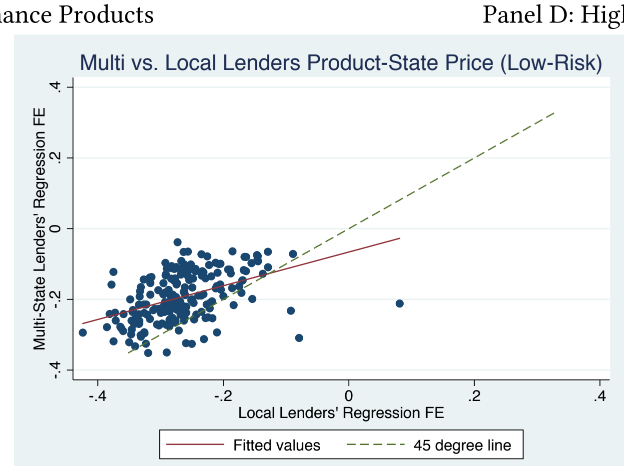
Panel B: Purchase Products



Panel C: Refinance Products



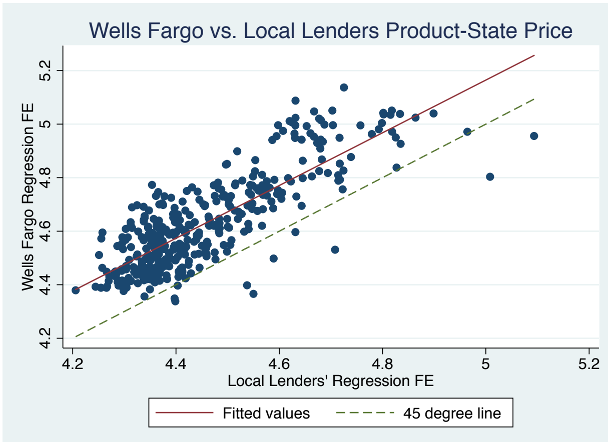
Panel D: High-Risk Products



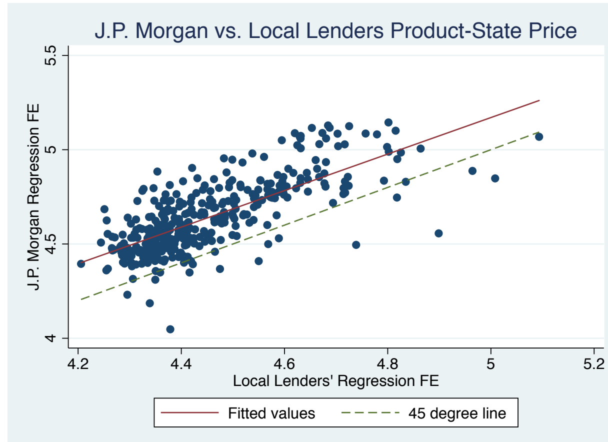
Panel E: Low-Risk Products

Figure 2.2: Multi-state vs. Local Lenders Product-State Fixed Effects

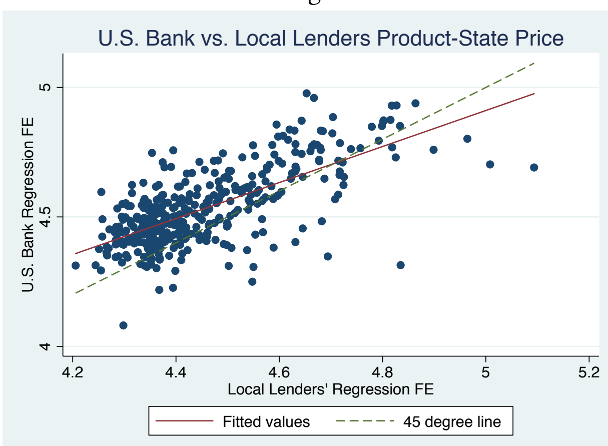
This figure presents the estimated local and multi-state lender fixed effects in equation 2.1 for each available product j, state s combination using GSE loans only. The local lender fixed effects α_{js}^{local} are plotted on the x-axis, while the multi-state lender fixed effects α_{js}^{multi} are plotted on the y-axis. The solid red line is the prediction for α_{js}^{multi} from a linear regression of α_{js}^{multi} on α_{js}^{local}. The green dotted line is the 45 degree line. Panel A contains all 8 products in our GSE sample. Panels B to E contain purchase, refinance, high-risk, low-risk products, respectively.



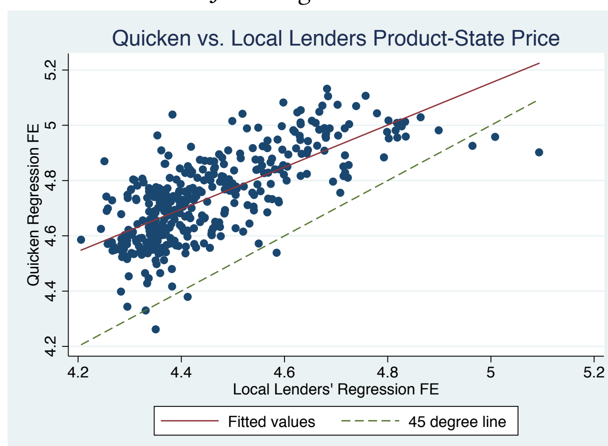
Panel A: Wells Fargo GSE 8 Products



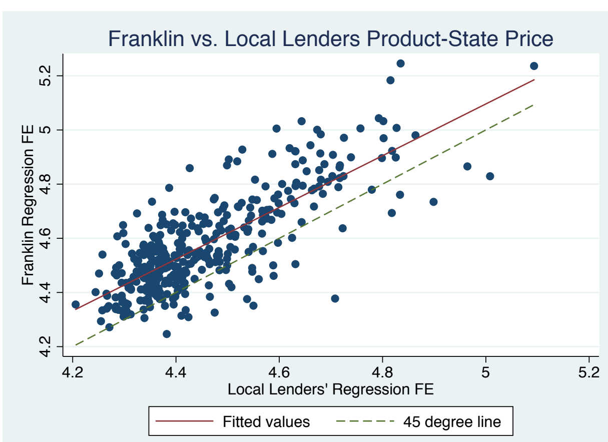
Panel B: J.P. Morgan GSE 8 Products



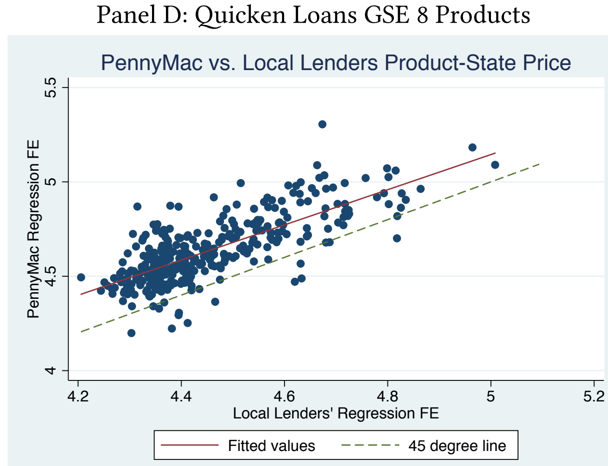
Panel C: U.S. Bank GSE 8 Products



Panel D: Quicken Loans GSE 8 Products



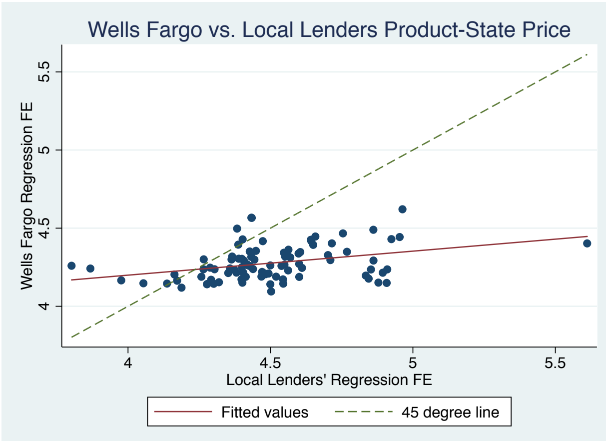
Panel E: Franklin American MTG Co. GSE 8 Products



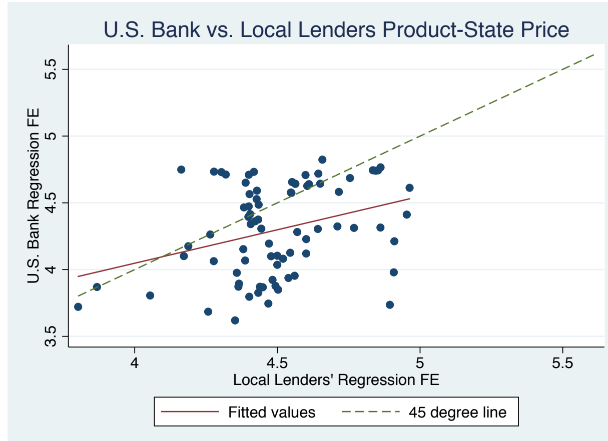
Panel F: PennyMac Co. GSE 8 Products

Figure 2.3: GSE Top 3 National Bank/Non-Bank Lender vs. Local Lenders Product-State FE

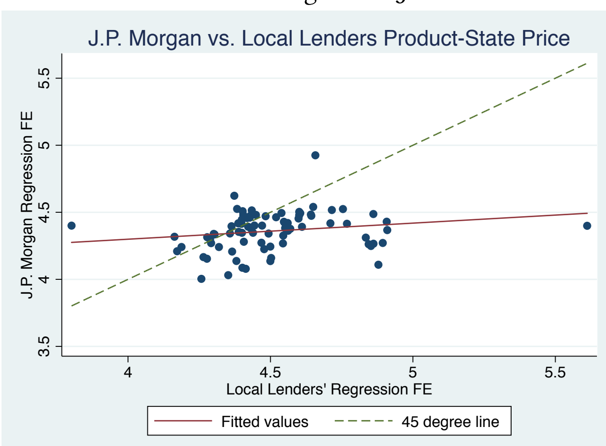
This figure depicts the relationship between $\alpha_{js}^{top_i}$ and α_{js}^{local} estimated in equation 2.2 using GSE loans. In panels A to C, we have the top 3 banks in the GSE loan market: Wells Fargo, J.P. Morgan, U.S. Bank. In panels D-F, we have the top 3 non-bank lenders in the GSE loan market: Quicken Loans, Franklin American Mortgage Company and PennyMac. Each point $(X, Y) = (\alpha_{js}^{local}, \alpha_{js}^{top_i})$ represents the estimated fixed effects in equation 2.2 for every product j, state s this top i lender is active in.



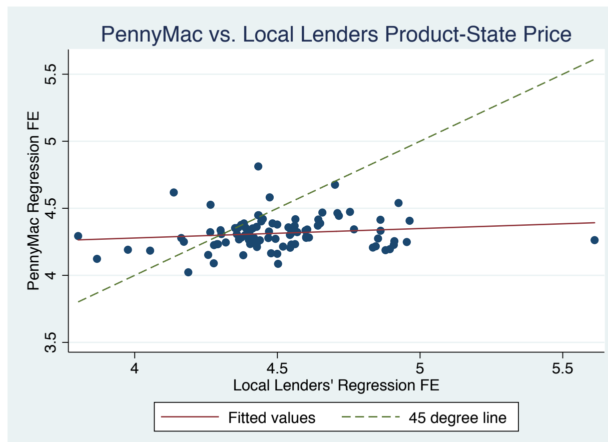
Panel A: Wells Fargo FHA 5 Products



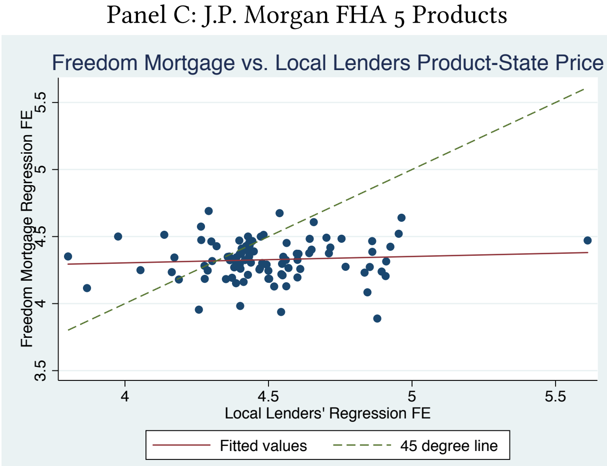
Panel B: U.S. Bank FHA 5 Products



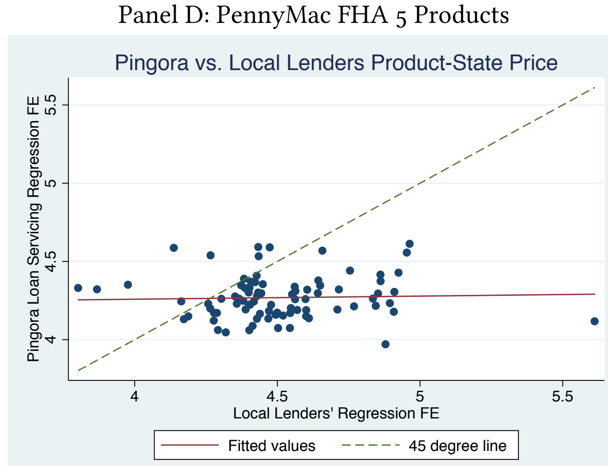
Panel C: J.P. Morgan FHA 5 Products



Panel D: PennyMac FHA 5 Products



Panel E: Freedom MTG Co. FHA 5 Products



Panel F: Pingora Loan Servicing FHA 5 Products

Figure 2.4: FHA Top 3 National Bank/Non-Bank Lender vs. Local Lenders Product-State FE

This figure depicts the relationship between $\alpha_{js}^{top_i}$ and α_{js}^{local} estimated in equation 2.2 using FHA loans. In panels A to C, we have the top 3 banks in the FHA loan market: Wells Fargo, U.S. Bank, and J.P. Morgan. In panels D-F, we have the top 3 non-bank lenders in the FHA loan market: PennyMac, Freedom Mortgage Company, and Pingora Loan Servicing. Each point $(X, Y) = (\alpha_{js}^{local}, \alpha_{js}^{top_i})$ represents the estimated fixed effects in equation 2.2 for every product j, state s this top i lender is active in.

Chapter 3

Increasing Price Asymmetry in the U.S. Retail Gasoline Market

3.1 Introduction

It has long been noticed by the public that oil prices seem to move asymmetrically. While retail gasoline prices respond almost instantaneously to crude oil price rises, they seem reluctant to move downwards when crude oil prices fall. This phenomenon has been observed in various different time periods and regions and has drawn the attention of not only many economic scholars, but also businessmen and politicians, leading to many worthy attempts to better understand this pricing pattern.

This phenomenon was first described by Bacon (1991) as "rockets and feathers", gasoline prices shooting up like rockets in response to positive oil price shocks and floating down like feathers in response to negative oil price shocks. Using a quadratic quantity adjustment function and semimonthly U.K. retail gasoline data from 1982 to 1989, Bacon reports evidence of a faster and more concentrated response of the retail price to spot price increases than decreases. Following this work, Borenstein, Cameron, and Gilbert (1997) propose in their seminal paper a symmetric error correction model (ECM) and study U.S. gasoline price transmission at different points of the distribution chain from 1986 to 1990. They find that spot prices for generic gasoline respond asymmetrically to crude oil price changes which may reflect production and inventory adjustment lags, and retail prices respond asymmetrically to wholesale price changes which may indicate short-run market power among retailers. However, they admit that the exact mechanism for this asymmetric pricing pattern is still unclear, since "lags in the adjustment of price to input cost changes are not consistent with simple models of either competitive markets or monopoly".

In addition, this asymmetric speed in pass-through for positive and negative shocks is not spe-

Earlier versions of this chapter have benefited from the comments of George Deltas, Andriy Norets, and Roger Koenker. All remaining errors and omissions are my own.

cific to the gasoline market. Peltzman (2000) analyzes 77 consumer and 165 producer goods in the U.S. and finds that output prices tend to respond faster to input price increases than to decreases in two of every three markets examined. He also establishes correlations between the degree of asymmetry and input price volatility, but finds no correlation with commonly mentioned reasons such as proxies for inventory costs, asymmetric menu costs of price changes, and imperfect competition.

In more recent years, the literature on asymmetric gasoline price pass-through rates mainly focused on two directions. The first is to explore how various data frequency, sample periods, geographic markets, and model specifications affect the existence and the degree of asymmetry in different segments of the industry. However, the results drawn from this line of research show great disparity, making it difficult to reach definitive conclusions. Borenstein et al.'s result was confirmed by Galeotti et al. (2003) using European data but rejected by Bachmeier and Griffin (2003) using U.S. daily data. Moreover, Bettendorf et al. (2003) find that the choice of the day when prices are observed greatly influences the results and conjecture this as a possible reason for the lack of robust conclusions in the literature. Deltas (2008) uses monthly data from 48 contiguous U.S. states and show that the degree of asymmetry in retail price adjustment increases with the average price-cost margin, which suggests a link between gas station market power and asymmetric responses. Perdiguero-Garciá (2013) provides a detailed meta analysis to unveil which specific features account for these differences and concludes with a likely relationship between the level of competition and the existence of asymmetries in the market. The second thread of literature focuses on developing theoretical models that could explain this asymmetric adjustment in gasoline pricing and providing corroborative empirical evidence. Potential reasons discussed include tacit collusion (Borenstein and Shepard, 1996), adjustment costs of production, costly inventory, market power (Borenstein and Shepard, 2002), consumer search (Johnson, 2002; Lewis, 2011), and oligopolistic coordination (Radchenko 2005).

In conclusion, due to the complicated structure and dynamics in the oil industry and gasoline market, the existence and degree of asymmetry is highly sensitive to data and model selections. Furthermore, the mechanism of asymmetric price adjustment is still not perfectly understood. However, having rigid prices and asymmetric adjustments is at least some indication of the existence of market power and related inefficiencies.

This chapter attempts to compare the degree of gasoline price asymmetry in the U.S. across time. Using weekly spot market and retail gasoline data from 1993 to 2013, I first confirm the existence of asymmetric price adjustments in this 20-year span with a generalized asymmetric error correction model. I then partition the data into different periods based on the structural breakpoints detected. Each period is analyzed individually before doing a comparison of the parameters estimated, the cumulative response functions, and consumer costs. Since the Great Recession caused significant oil price decreases in 2008, I also discarded data from this abnormal period and repeated the aforementioned analysis. In addition, I use a time-varying coefficient model to estimate the previous coefficients as functions of time. This allows me to compare the retail gasoline price responses to spot market price increases or decreases at different points in time. While different results were drawn from a wide variety of model specifications, data sets and regions across numerous studies in the literature, I hold these factors constant and change only the time period of the study for a more reliable comparison. My results indicate that retail gasoline price movements have become increasingly asymmetric in recent years, leading to inefficiencies such as larger consumer costs.

3.2 U.S. Gasoline Industry Structure

Gasoline is made from crude oil and other petroleum liquids and mainly used as an engine fuel in vehicles. After gasoline is produced at U.S. refineries, it is usually transported through pipelines, tankers, or barges to terminals that provide storage and dispensing facilities. From terminals, trucks bring gasoline products to retail outlets through wholesale distribution networks, and these retail stations eventually deliver gasoline to the consumer.

Refiners often sell large quantities of generic gasoline from the refinery to distributors or other refiners in spot transactions. Generic gasoline prices are reflected in the spot gasoline prices for delivery to New York Harbor and the Gulf Coast, which are highly correlated.[1] Branded refineries, Shell, Chevron, Exxon, etc., ship their gasoline to the distribution terminal in a city where it is sold as either branded gasoline or generic gasoline. Branded retail stations are required to purchase

[1]Similar to crude oil, gasoline spot prices are determined by a daily survey of major traders. I use the New York Harbor spot price in this chapter.

that brand of fuel, which usually contains company specific additives. Unbranded refineries do not operate their own chain of retail outlets and sell unbranded gasoline at their city terminals for resale at unbranded stations. From the city terminal, gasoline might be distributed directly by the refiner or through middlemen know as jobbers, who typically supply multiple stations of different brands and generally owns many of the stations it supplies. A large percentage of U.S. gasoline is distributed by jobbers or through other companies that are not controlled by refiners. The rest is transported from the terminal to the retailer by the refiner. These direct-supplied stations are operated by an independent retailer, lessee dealer or the refiner. See Figure 3.1 for the physical structure of gasoline distribution and marketing channels in the U.S.

Gasoline prices are observed at various points of the transmission from the refinery to the service station, including regional spot markets, wholesale city terminals, dealer tank wagons, and retail pumps. Borenstein, Cameron, and Gilbert (1997) find immediate responses in wholesale prices to spot market price changes and suggest two noteworthy asymmetric adjustments, the pass-through of crude oil price changes to the spot market price and the pass-through of spot market price changes to the retail price. Perdiguero-Garciá (2013) points out that asymmetric behavior is more likely to be found in the retail segment, that is, between crude oil, spot prices or wholesale prices and the retail price paid by consumers. Potentially, this is because retail is the part of the transmission where the market is more concentrated and less competitive. Following this direction, this chapter focuses on how positive and negative spot market price shocks are reflected in the retail gasoline price available to consumers.

3.3 Data

I use weekly New York Harbor (NYH) conventional gasoline regular spot price FOB (Dollars per Gallon) and weekly U.S. regular conventional area retail gasoline prices (Dollars per Gallon) for spot market and retail gasoline prices, respectively.[2][3] This data is provided by the Energy

[2] Conventional gasoline is not included in the reformulated gasoline category and excludes reformulated gasoline blendstock for oxygenate blending (RBOB) as well as other blendstock. Free on board (FOB) pertains to a transaction whereby the seller makes the product available within an agreed on period at a given port at a given price, and it is the responsibility of the buyer to arrange for the transportation and insurance. The results of this chapter remain unchanged with the Gulf Coast spot prices, which is highly correlated to NYH spot prices.

[3] A conventional area does not require the sale of reformulated gasoline and hence, all types of finished motor gasoline may be sold in this area.

Information Administration (EIA) and the time period covered in this analysis is from April 1993 to April 2013.

EIA reports retail gasoline price every Monday while the NYH spot market price is reported daily. Every Monday, retail prices for all three grades of gasoline (regular, midgrade, and premium) are collected by telephone, email, text, fax, or the internet from a sample of gasoline outlets across the U.S. using the EIA's Form 878, "Motor Gasoline Price Survey". This survey is designed to collect and publish data on the cash price (including taxes) of self-serve, unleaded gasoline, by grade of gasoline. The sample was drawn from a frame of approximately 115,000 retail gasoline outlets.[4] The prices are published around 5:00 p.m. ET Monday, except on government holidays, when the data is released on Tuesday (but still represent Monday's price). The reported price includes all taxes and is the pump price paid by the consumer as of 8:00 a.m. on Monday for self-serve stations, except in areas that have full-serve only. The price data is used to calculate weighted average price estimates at the city, state, regional, and national levels using sales and delivery volume data from other EIA surveys and population estimates from the Census Bureau.

In this chapter, I match every Monday's retail price with the previous week's NYH spot market price, where weekly NYH spot prices are calculated by the EIA from daily data by taking an unweighted average of the daily closing spot prices for a given product over the specified time period. Perdiguero-Garciá (2013) suggests that more frequent data increases the probability of detecting asymmetric price responses. Hence, I choose to use national level data, which is available weekly starting from 1993, while state and city level data are only available monthly and start later in 2000 or 2003.

3.4 Econometric Model

The purpose of this chapter is to examine the transmission of spot market price shocks to retail gasoline prices. Hence, my econometric model specification abstracts from other determinants of the retail gasoline price such as inventory levels, station capacity, future gasoline price pre-

[4] More information on the EIA's gasoline price data collection procedure including sampling methods, imputation and estimation, sampling errors, and nonsampling errors can be found at

dictions, and treats the retail gasoline price as an autoregressive process which depends on a distributed lag of current and past spot market prices.

I follow the setup of Borenstein et al. (1997) and analyze in first differences instead of in levels to ease the comparison of my results with other papers. The generalized asymmetric error correction model I use is the following:

$$\Delta r_t = \Sigma_{i=0}^{k}\beta_{si}^{+}\Delta s_{t-i}^{+} + \Sigma_{i=1}^{m}\beta_{ri}^{+}\Delta r_{t-i}^{+} + \theta^{+}\hat{e}_{t-1}^{+} + \Sigma_{i=0}^{k}\beta_{si}^{-}\Delta s_{t-i}^{-} + \Sigma_{i=1}^{m}\beta_{ri}^{-}\Delta r_{t-i}^{-} + \theta^{-}\hat{e}_{t-1}^{-} + \varepsilon_t \tag{3.1}$$

where Δr_t and Δs_t are first differences of retail and spot market gasoline prices, respectively. To differentiate between positive and negative changes in these price series, I denote

$$\Delta s_{t-i}^{+} = \max\{\Delta s_{t-i}, 0\}; \quad \Delta s_{t-i}^{-} = \min\{\Delta s_{t-i}, 0\}$$

$$\Delta r_{t-i}^{+} = \max\{\Delta r_{t-i}, 0\}; \quad \Delta r_{t-i}^{-} = \min\{\Delta r_{t-i}, 0\}$$

$$\hat{e}_{t-1}^{+} = \max\{r_{t-1} - \hat{\gamma}_0 - \hat{\gamma}_1 s_{t-1}, 0\}; \quad \hat{e}_{t-1}^{-} = \min\{r_{t-1} - \hat{\gamma}_0 - \hat{\gamma}_1 s_{t-1}, 0\}$$

The coefficient β_{si}^{+} (β_{si}^{-}) measures the short-run impact of period $(t-i)$ spot market price increases (decreases), while β_{ri}^{+} (β_{ri}^{-}) measures the short-run impact of positive (negative) period $(t-i)$ retail gasoline prices. The coefficients θ^{+} and θ^{-} are the long-run equilibrium adjustment parameters. The long-run positive (negative) disequilibrium between spot and retail gasoline prices is represented by $\hat{e}_{t-1}^{+}$ ($\hat{e}_{t-1}^{-}$), which follows a stationary process because spot and retail gasoline prices are conintegrated. As shown in Engle and Granger (1987), the cointegrating vector parameters $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are superconsistent OLS estimates, where the long-run effect of a permanent change in spot market gasoline prices is $\hat{\gamma}_1$. Although asymmetric price adjustments occur in the short run, the error correction model assumes the underlying long run equilibrium relationship between spot market and retail gasoline prices is the same when price increases or decreases. When using the whole sample, the number of lags k and m, are chosen to be 1 and 3 by the Schwarz information criterion (1978), respectively.

3.5 Estimation Results

3.5.1 Whole Sample Period Results

I start the analysis with the entire sample period from April 1993 to April 2013. Estimation results of the asymmetric ECM using this 20-year period are shown in Table 3.1 column (1). The results indicate that the contemporaneous response of retail prices to spot market gasoline price increases β^{+}_{s0} is larger than the response to spot market gasoline price decreases β^{-}_{s0} and the symmetric specification is rejected according to the Wald test. Yet the retail prices' response to spot market gasoline price changes in the previous period has an opposite effect, namely, the response to positive changes β^{+}_{s1} is smaller than the response to negative changes β^{-}_{s1}, albeit the magnitude being much smaller than the current period responses. Consistent with previous studies, the coefficient for spot market prices in the cointegration relationship $\hat{\gamma}_1$ is approximately 1, suggesting a one-for-one cost to price pass-through in the long-run. To fully measure the adjustment path of retail prices to a one-unit change in upstream NYH spot prices, I construct cumulative adjustment functions for increases in spot prices with the following formula:

$$
\begin{aligned}
B^{+}_{0} &= 0 \\
B^{+}_{1} &= \beta^{+}_{s0} \\
B^{+}_{2} &= B^{+}_{1} + \beta^{+}_{s1} + \theta^{+} max(B^{+}_{1} - \gamma_1, 0) + \theta^{-} min(B^{+}_{1} - \gamma_1, 0) \\
&+ \beta^{+}_{r1} max(B^{+}_{1} - B^{+}_{0}, 0) + \beta^{-}_{r1} min(B^{+}_{1} - B^{+}_{0}, 0) \\
\vdots &= \vdots \\
B^{+}_{k} &= B^{+}_{k-1} + \beta^{+}_{s,k-1} + \theta^{+} max(B^{+}_{k-1} - \gamma_1, 0) + \theta^{-} min(B^{+}_{k-1} - \gamma_1, 0) \\
&\quad + \Sigma^{k-1}_{i=1} [\beta^{+}_{ri} max(B^{+}_{k-i} - B^{+}_{k-i-1}, 0) + \beta^{-}_{ri} min(B^{+}_{k-i} - B^{+}_{k-i-1}, 0)]
\end{aligned}
$$

where B^{+}_{k} is the cumulative adjustment function of retail prices in period k for an initial one-unit increase in upstream NYH spot prices. These functions capture both the indirect effects from lagged changes in retail prices and the effect of the long-run relationship reversion. Cumulative adjustment functions for an initial decrease are defined similarly and standard errors are derived using bootstrap methods.[5]

The estimated cumulative adjustment functions are shown in Figure 3.2 Panel A. The yellow line with circled markers is the estimated retail price response in dollars per gallon to a one-time \$1 per gallon increase in the spot market gasoline price. A \$1 increase in the spot market gasoline price leads to a \$0.58 increase in the first week, a further \$0.15 increase in the second week, approximately \$0.9 increase after five weeks, and so on. The green line with triangle markers is the estimated retail price response to a decrease in spot market gasoline prices. The cumulative response functions drawn in the graphs are actually the negative value of the cumulative response functions defined above. This allows the functions to converge to 1 when the pass-through rate is 100% instead of -1, which makes it easier to compare the speed of convergence and consumer costs (defined later as the area between the two cumulative adjustment functions). While the retail price goes up immediately when there is a \$1 increase in the spot market price, the retail price continues to rise for the first couple of weeks before it starts to go down when there is a \$1 decrease in the spot market price. The former series converges to 1 after 9 weeks while the latter converges to 1 after 15 weeks or so. The dashed lines are the estimates of the 95% confidence intervals for the cumulative adjustment functions using bootstrap methods described earlier. As expected, estimates of these cumulative functions get noisier when it is further away from the date of the one-time price change. Still, the functions are sufficiently different to signify asymmetric adjustment speeds.

Following Borenstein, Cameron and Gilbert (1997), the adverse consequences of asymmetric pricing can be evaluated in terms of consumer costs. I now measure the welfare effects of asymmetry by comparing the consumer gains from a \$1 decrease in the spot market oil price over the lifetime of the price adjustment process to the consumer losses over the adjustment process from an equal size increase in the spot price. Integrating the differences in cumulative adjustments yields an estimate of the consumer costs:

$$\Delta Consumer\ Cost = \int_{j=0}^{n} (B_j^+ - B_j^-)dj \tag{3.2}$$

[5]The bootstrap is conducted by first drawing from a multivariate normal distribution with means equal to the estimated coefficients and covariance equal to the estimated covariance matrix, and then calculating the price responses over time for each draw (500 draws were taken for each time point). The confidence bands are formed by connecting the point-wise 95% confidence intervals for each time point after the initial price change.

where B_j^+ and B_j^- are the estimated cumulative adjustments at time j to a \$1 increase and a \$1 decrease in spot market gasoline prices, respectively. Under simple linear interpolation between estimated adjustment points, consumer cost is the difference in the areas under the two cumulative adjustment curves from week 0 to week n. The fact that retail prices react slower to spot market price decreases leads to net costs for consumers. As shown in Figure 3.2 Panel B, volatility even without net changes in spot market prices is costly for consumers and the consumer costs converge to approximately \$7 after 15 weeks. Estimates of the 95% confidence intervals for the cumulative consumer costs are once again bootstrapped.

However, obvious changes in both level and volatility for spot and retail gasoline prices (see Figure 3.3) cast doubt on the validity of analyzing the full sample period as a whole. In addition, the goal of this chapter is to detect potential patterns in the degree of asymmetry responses across time. Therefore, in the following two subsections, we divide our data into different periods and conduct analyses for each period separately.

3.5.2 Structural Break Periods

Using the algorithm described in Bai & Perron (2003) for simultaneous estimation of multiple breakpoints, I find significant evidence for structure breaks in the retail gasoline price and NYH spot gasoline price time series.[6] In Figure 3.3 Panel A, the original retail price series is plotted alongside the fitted values of the regression $r = \alpha + \beta t$, where r is the weekly retail price and t is the number of weeks. The green dashed line represents the fitted values without breakpoints while the blue dotted lines are the fitted retail prices with the number of breakpoints selected by the Bayesian information criterion (BIC). The four detected breakpoints are January 9, 1998, September 21, 2001, February 25, 2005, and October 10, 2008. The red intervals enclosing each breakpoint are the 95% confidence intervals. Similarly, in Figure 3.3 Panel B, I plot the NYH spot price series with fitted values of the regression $s = \alpha + \beta t$, where s is the weekly spot price and t is the number of weeks. The detected breakpoints are December 26, 1997, September 14, 2001, September 30, 2005, and September 26, 2008. Note that except for the third breakpoint, the other three breakpoints for these two price series are actually very close. In this section, I partition

[6]The distribution function used for the confidence intervals for the breakpoints is given in Bai (1997). The ideas behind this implementation are described in Zeileis et al. (2003).

the data with the detected breakpoints for retail prices. Dividing the data according to the four breakpoints for NYH spot prices gives similar results.

According to the observable price dynamics in our sample, for both retail and NYH spot prices, the first period is the most stable. The second and third periods exhibit more price volatility and have positive time trends with similar slopes. The fourth period has the most unstable prices and a much larger slope indicating retail gasoline prices rapidly increasing with time. The fifth period includes the huge price drop from July 2008 to late November 2008, which is mainly due to the global recession's significant impact on oil prices during this time. It also includes post-crisis data from late 2008 to April 2013, during which oil prices bounced back to approximately its pre-crisis level.

We now compare the estimation results for the five periods above to see whether there is asymmetric price adjustment behavior in each period and how the degree of asymmetry changes across time. The regression results of the asymmetric ECM using data from these five periods are shown in columns (2)-(6) of Table 3.1. Note that the number of lags k and m, are now both chosen to be 1 by the Schwarz information criterion. We observe that the contemporaneous response of retail prices to spot market gasoline price increases β_{s0}^{+} is larger than the response to spot market gasoline price decreases β_{s0}^{-} for all five periods, indicating asymmetry in the initial retail price adjustments. In Figure 3.4 and Figure 3.5, the cumulative adjustment functions and consumer costs are plotted for the five periods. The comparison of consumer costs across five periods is shown in Figure 3.6. For all five periods, similar to the whole sample, I again observe the retail price going up immediately when there is a $1 increase in the spot market price while the retail price continues to rise for the first couple of weeks before it starts to go down when there is a $1 decrease in the spot market price. Among the five periods, the response to a $1 decrease catches up most quickly and surpasses that of a $1 increase in period 2 at around week 7, causing consumer costs to decrease from week 7 and statistically indistinguishable from zero after week 9. For all other periods, consumer costs remain significantly different from zero even after 15 weeks. Moreover, the first four periods have relatively similar consumer costs (2 to 3 dollars) while in the last period consumers apparently suffer from larger costs (6 dollars) due to asymmetric price responses.

3.5.3 Discarding Irregular Data from the Great Recession Period

In this section, the data set is divided into three time periods, as presented in Figure 3.7. The most obvious breakpoint is in 2008 when the economic crisis greatly influenced the oil industry. The U.S. average price for regular gasoline climbed to an all-time high of $4.11 per gallon in July 7, 2008 and then plummeted to a 5-year low within only a couple of months. We discarded data from August 8, 2008 to March 27, 2009, the period when gasoline prices fell dramatically. Moreover, including data from this period causes the series to fail the premise of an error correction model- the Augmented Dickey-Fuller unit root test. Therefore, the last period starts from April 3, 2009 and ends at March 29, 2013. The first two periods are defined due to their differences in time trend and volatility. The first period, from April 2, 1993 to March 16, 2001, has relatively steady prices, in terms of both level and volatility. The second period, from March 23, 2001 to August 1, 2008, displays a more obvious positive trend and larger volatility.

The regression results for the asymmetric ECM using data from these three periods are reported in Table 3.2 columns (1)-(3), respectively. The number of lags k and m are again both chosen to be 1. The contemporaneous response of retail prices to spot market gasoline price increases β^{+}_{so} is larger than the response to spot market gasoline price decreases β^{-}_{so} and the symmetric specification is rejected according to the Wald test for all three periods. The estimated cumulative adjustment functions and consumer costs for periods 1-3 are shown in Figure 3.8 and the comparison across three periods is shown in Figure 3.9. In period 1, after approximately week 9, the response to a $1 decrease surpasses that of a $1 increase. Apart from that, increases always seem to be passed along faster than decreases. Since retail prices react slower to spot market price decreases, price adjustment asymmetry leads to net costs for consumers. In period 1, consumer cost rises to $2.65 at around week 10 before it starts to fall and is larger than zero before week 14 at the 95% significance level. In period 2 and period 3, consumer cost converges to approximately $5.5 and $6, respectively. Figure 3.9 shows an comparison across periods where consumer costs for the first period is significantly smaller than the latter two periods.[7][8]

[7]The prices here are nominal. However, the trend of growing consumer costs still hold for real prices. The cumulative rate of inflation is 19.2% for 1993-2000, 49% for 1993-2008 and 61.2% for 1993-2013.

[8]Estimates of the 95% confidence intervals for the cumulative adjustment functions are once again obtained using bootstrap methods.

3.6 Time-Varying Coefficient Model

In the previous sections, I separated the data based on different price dynamics and structural break test results. In this section, the data is pooled in a time-varying coefficient model. By obtaining the estimates in a single nested model and as functions of time, smoother changes in the estimated parameters are allowed. Using BIC as our model selection criteria, the optimal model with pooled data and constant coefficients is the following:

$$\Delta r_t = \Sigma_{i=0}^{1}\beta_{si}^{+}\Delta s_{t-i}^{+} + \Sigma_{i=1}^{3}\beta_{ri}^{+}\Delta r_{t-i}^{+} + \theta^{+}\hat{e}_{t-1}^{+} + \Sigma_{i=0}^{1}\beta_{si}^{-}\Delta s_{t-i}^{-} + \Sigma_{i=1}^{3}\beta_{ri}^{-}\Delta r_{t-i}^{-} + \theta^{-}\hat{e}_{t-1}^{-} + \varepsilon_t \tag{3.3}$$

My time-varying coefficient model maintains this specification, including the number of lags selected, while allowing all the coefficients to be functions of time:

$$\begin{aligned} \Delta r_t &= \Sigma_{i=0}^{1}\beta(t)_{si}^{+}\Delta s_{t-i}^{+} + \Sigma_{i=1}^{3}\beta(t)_{ri}^{+}\Delta r_{t-i}^{+} + \theta(t)^{+}\hat{e}_{t-1}^{+} \\ &+ \Sigma_{i=0}^{1}\beta(t)_{si}^{-}\Delta s_{t-i}^{-} + \Sigma_{i=1}^{3}\beta(t)_{ri}^{-}\Delta r_{t-i}^{-} + \theta(t)^{-}\hat{e}_{t-1}^{-} + \varepsilon_t \end{aligned} \tag{3.4}$$

The plots for functions $\beta_{s0}^{+}(t)$, $\beta_{s0}^{-}(t)$, $\beta_{s1}^{+}(t)$ and $\beta_{s1}^{-}(t)$ are shown in Figure 3.10 Panel A. On one hand, the estimated contemporaneous response of retail gasoline prices to spot market price increases $\beta_{s0}^{+}(t)$ is larger than the response to spot market gasoline price decreases $\beta_{s0}^{-}(t)$ during the entire sample period. On the other hand, the estimated response of a previous period spot market gasoline price increase $\beta_{s1}^{+}(t)$ is less than the estimated response of a previous period spot market gasoline price decrease $\beta_{s1}^{-}(t)$. However, the latter two coefficient functions are smaller in magnitude for the entire sample period. To have a rough estimate of the accumulated effect, the combination effect of two periods after a spot market price change is depicted in Figure 3.10 Panel B, with the accumulated response of a positive change always dominating a negative one, implying the pass-through of a positive shock being faster than a negative shock. The difference between the two functions first decreases and then increases. However, because the estimated coefficients are functions of time, after a one-unit change in the spot market price, instead of having the same cumulative response function for each period as in previous sections, here the cumulative response functions differ by the time the spot market price change occurred, making it relatively difficult to obtain a clean comparison across the entire sample period. All the estimated coefficient functions of this model are plotted in Figure 3.11.

3.7 Robustness Tests

In this section, I discuss the robustness of our empirical results. These robustness tests address some of the specifications that our model and analysis might be sensitive to and potential reasons mentioned in previous papers that might lead to inconclusive results. As a summary, none of the robustness specifications I explored materially change the conclusions of this chapter.

First, I add time and seasonal dummies to the asymmetric ECM to capture time and seasonal effects, the cumulative adjustment functions are hardly changed with these additional controls. Second, I estimate the model in one stage, as in Borenstein, Cameron and Gilbert (1997), instead of estimating the long-run relationship between retail and spot market price in the first stage and using the residuals as the error correction term in the second stage. The estimates obtained are very similar. Third, I change the corresponding NYH spot price for each week's reported retail price from the previous Friday to the next Monday, the results show that our conclusions are not affected by this change in the day of the week selected.

Additionally, to ensure that the results are not reliant on the particular division of data selected according to the structural breakpoints for $r = \alpha + \beta t$, I also repeat the analysis with the breakpoints detected by 1) regressing the NYH spot prices on time: $s = \alpha + \beta t$; 2) regressing the error terms obtained from the regression $r = \alpha + \beta t$ on a constant, i.e., structural breaks for the estimated error terms; 3) regressing the error terms obtained from the regression $s = \alpha + \beta t$ on a constant; 4) regressing retail prices on NYH spot prices, that is, structural breaks in their long-term relationship. The results for the first three partitions are mostly identical with our previous results in section 4.2 and therefore are omitted for brevity. The breakpoints selected for the last specification has a larger difference, as reported in Figure 3.12. The green dashed line is the fitted retail prices with a single long-term regression while the blue dotted line is the fitted retail prices with the number of breakpoints selected by BIC. The three detected breakpoints are March 10, 2000, September 2, 2005, and November 21, 2008. Structural breaks in the long-term relationship could be affected by various factors including market structure, production technology, transportation efficiency, customers' search behavior, etc. Although the break dates have changed, the pattern of increasing consumer costs due to asymmetric price adjustments remains the same.

3.8 Conclusion

In this chapter, I estimate the dynamic relationship between spot market and retail gasoline prices with a generalized asymmetric error correction model and a time-varying coefficient model using weekly data from April 1993 to April 2013. I test the existence of asymmetric price adjustments in the U.S. gasoline market with the entire 20-year sample and then partition the data according to structural break tests for comparison across time. Estimation results not only indicate that retail gasoline prices respond asymmetrically to spot market gasoline prices, but also the degree of asymmetry increasing across time. Previous research have linked this asymmetry in pass-through to various sources, most predominately, market power. Hence, the results of this chapter might be a potential signal of growing retail gasoline market power in the U.S. throughout the years of study.

3.9 Tables and Figures

Table 3.1: **Asymmetric ECM for Entire Sample and Five Structural Break Periods**

Regressor	(1) Entire Sample	(2) Period 1	(3) Period 2	(4) Period 3	(5) Period 4	(6) Period 5
Δr^{+}_{t-1}	0.0743*	0.280***	0.385***	0.291***	0.167**	0.182***
	(0.0375)	(0.0588)	(0.0640)	(0.0589)	(0.0556)	(0.0497)
Δr^{-}_{t-1}	0.239***	0.549***	0.342**	0.267**	0.244**	0.510***
	(0.0483)	(0.0954)	(0.113)	(0.0869)	(0.0905)	(0.0504)
Δr^{+}_{t-2}	0.0946***					
	(0.0240)					
Δr^{-}_{t-2}	0.0846*					
	(0.0412)					
Δr^{+}_{t-3}	0.0672**					
	(0.0235)					
Δr^{-}_{t-3}	0.0764*					
	(0.0335)					
Δs^{+}_{t}	0.583***	0.303***	0.488***	0.534***	0.598***	0.572***
	(0.0158)	(0.0278)	(0.0484)	(0.0405)	(0.0288)	(0.0362)
Δs^{-}_{t}	0.387***	0.135***	0.240***	0.331***	0.441***	0.373***
	(0.0176)	(0.0293)	(0.0472)	(0.0367)	(0.0409)	(0.0343)
Δs^{+}_{t-1}	0.0799**					
	(0.0271)					
Δs^{-}_{t-1}	0.1103**					
	(0.0242)					
$\hat{e}^{+}_{t-1}$	-0.116***	-0.0678***	-0.192***	-0.230***	-0.264***	-0.0918*
	(0.0186)	(0.0203)	(0.0398)	(0.0444)	(0.0583)	(0.0394)
$\hat{e}^{-}_{t-1}$	-0.0562***	-0.0621**	-0.0193	-0.106*	-0.138*	-0.141***
	(0.0164)	(0.0197)	(0.0450)	(0.0448)	(0.0541)	(0.0370)
Cointegration Relationship						
s_{t-1}	1.042***	0.812 ***	1.042***	1.053***	1.029***	0.966***
	(0.003)	(0.030)	(0.018)	(0.016)	(0.014)	(0.009)
constant	0.576***	0.677 ***	0.575 ***	0.571***	0.660***	0.735***
	(0.004)	(0.017)	(0.012)	(0.015)	(0.030)	(0.022)
Observations	1,040	248	193	179	189	233

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table reports the estimation results for the asymmetric error correction model using the entire 20-year sample from April 1993 to April 2013 in Column (1) and the five periods determined by the structural breakpoints detected in Columns (2)-(6), respectively. The number of lags were determined by the Schwarz information criterion. The long run equilibrium between spot market and retail gasoline prices in each period is characterized by the cointegration parameters shown in the second part of the table.

Table 3.2: **Estimation Results for Three Periods**

Regressor	(1) Period 1	(2) Period 2	(3) Period 3
Δr^+_{t-1}	0.337^{***}	0.235^{***}	0.204^{***}
	(0.0441)	(0.0339)	(0.0514)
Δr^-_{t-1}	0.311^{***}	0.372^{***}	0.467^{***}
	(0.0840)	(0.0561)	(0.0879)
Δs^+_t	0.399^{***}	0.582^{***}	0.556^{***}
	(0.0273)	(0.0179)	(0.0377)
Δs^-_t	0.201^{***}	0.434^{***}	0.351^{***}
	(0.0274)	(0.0248)	(0.0363)
$\hat{e}^+_{t-1}$	-0.146^{***}	-0.141^{***}	-0.102^{*}
	(0.0235)	(0.0317)	(0.0453)
$\hat{e}^-_{t-1}$	-0.0443^{*}	-0.115^{***}	-0.131^{***}
	(0.0217)	(0.0310)	(0.0378)
Cointegration Relationship			
s_{t-1}	0.986^{***}	1.056^{***}	0.963^{***}
	(0.013)	(0.006)	(0.012)
constant	0.588^{***}	0.586^{***}	0.742^{***}
	(0.008)	(0.009)	(0.031)
Observations	414	385	209

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table presents the asymmetric error correction model estimation results for the three periods discussed in Section 5.3. The number of lags were determined by the Schwarz information criterion. The long run equilibrium between spot market and retail gasoline prices in each period is characterized by the cointegration parameters shown in the second part of the table.

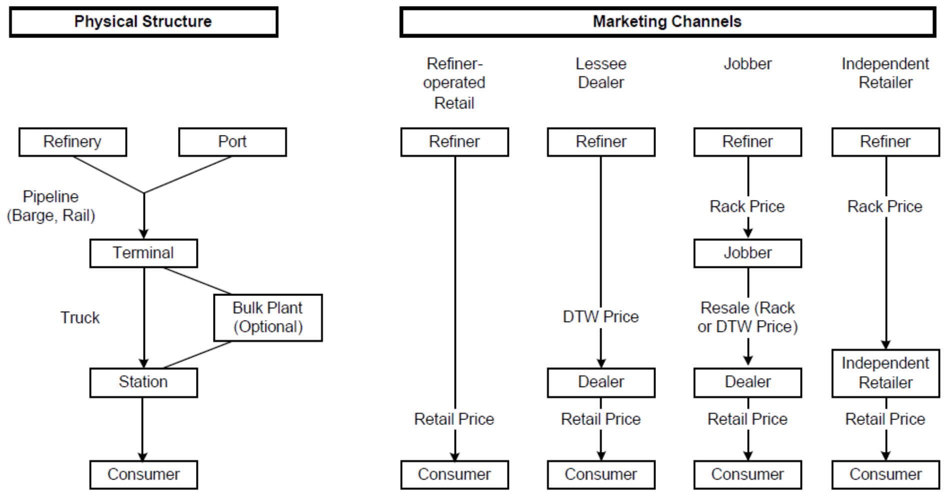
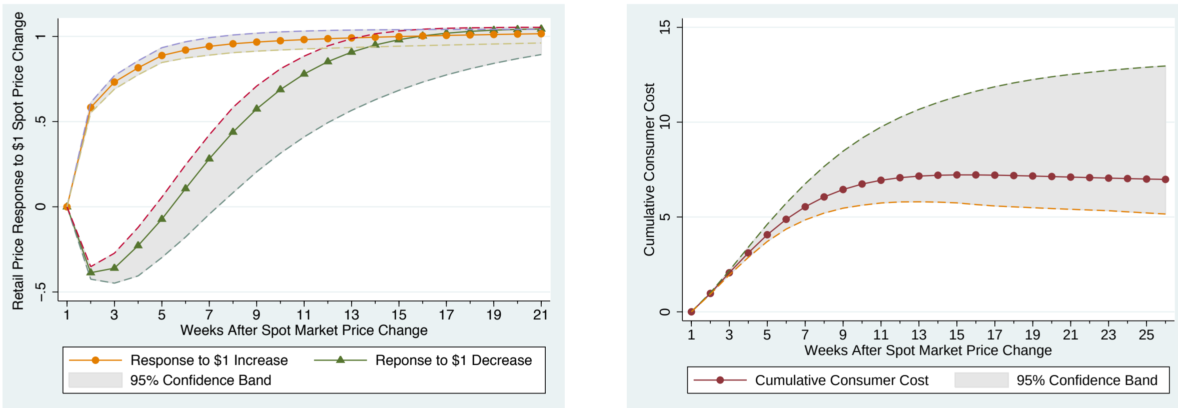


Figure 3.1: Gasoline Distribution Physical Structure and Marketing Channels

Source: U.S. Department of Energy, Energy Information Administration (EIA).

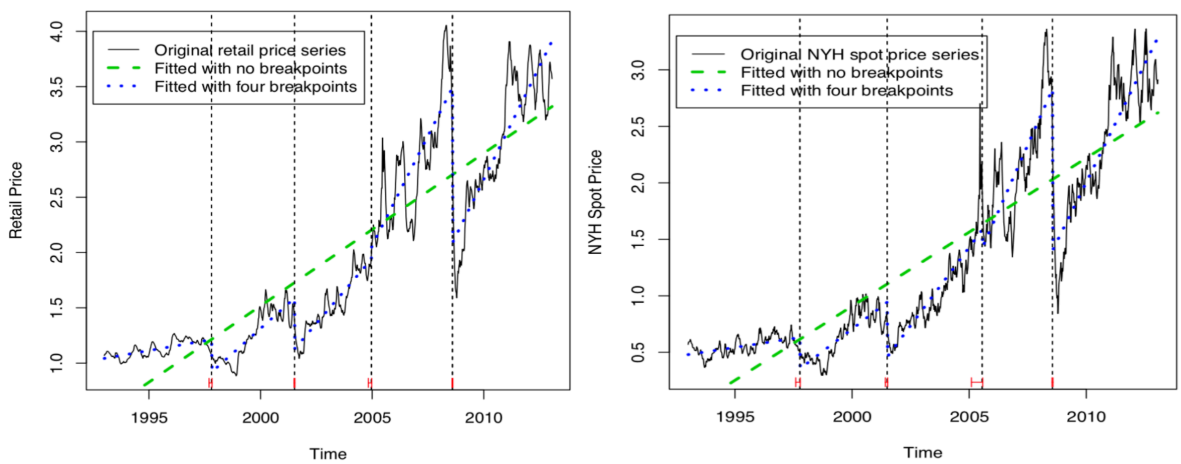


Panel A: Cumulative Adjustment Function

Panel B: Cumulative Consumer Costs

Figure 3.2: Cumulative Adjustment Function and Consumer Costs (April 1993-April 2013)

This figure depicts the cumulative adjustment function after a one-unit change in the spot market price (Panel A) and the cumulative consumer costs due to asymmetric price adjustments (Panel B) calculated with the entire 20-year sample. The 95% confidence intervals are obtained with bootstrapping methods.

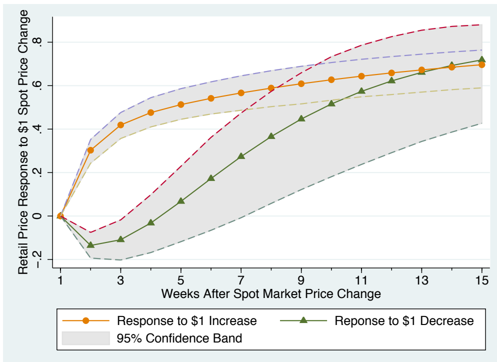


Panel A: Structural Breakpoints for Retail Prices

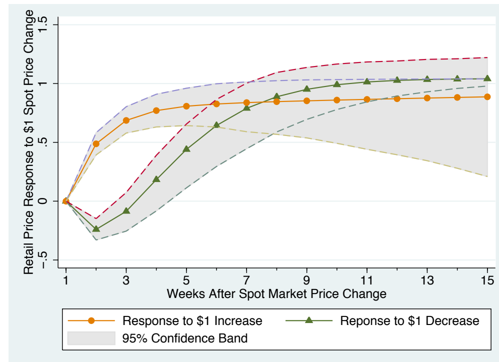
Panel B: Structural Breakpoints for NYH Spot Prices

Figure 3.3: Structural Breaks for Retail and NYH Spot Gasoline Prices

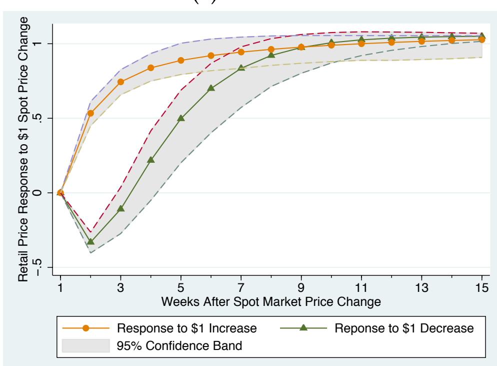
This figure presents the detected structural breakpoints for the retail (Panel A) and spot market (Panel B) price series and fitted values with (blue dotted lines) and without (green dashed line) structural breaks. In both panels, the original price series is plotted alongside the fitted values of the without breakpoints. The four detected breakpoints for the retail gasoline prices are January 9, 1998, September 21, 2001, February 25, 2005, and October 10, 2008. The detected breakpoints for the spot market gasoline prices are December 26, 1997, September 14, 2001, September 30, 2005, and September 26, 2008.The red intervals enclosing each breakpoint are the 95% confidence intervals.



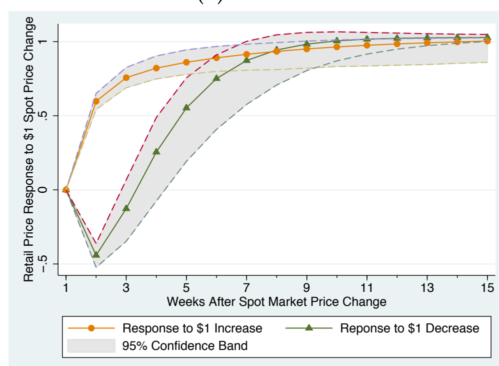
(a) Period 1



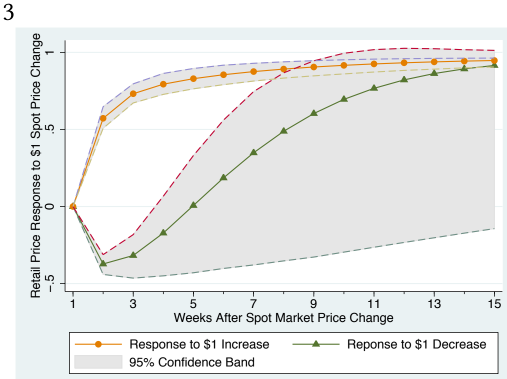
(b) Period 2



(c) Period 3



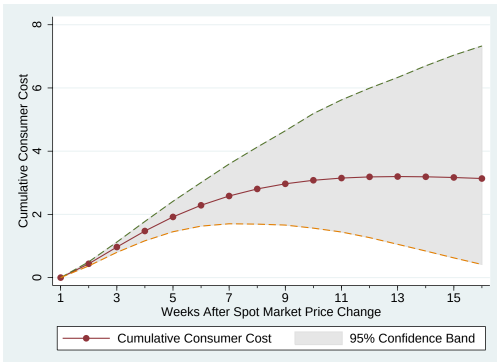
(d) Period 4



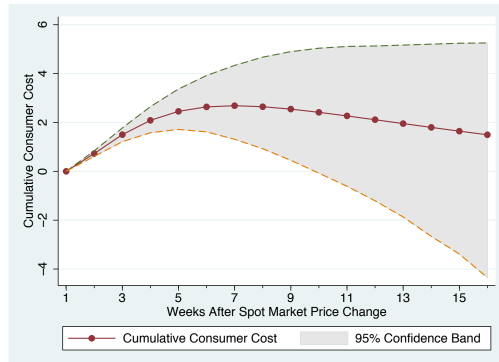
(e) Period 5

Figure 3.4: Cumulative Adjustment Functions for Five Structural Break Periods

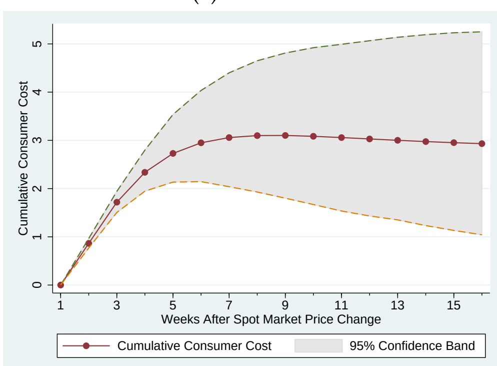
This figure illustrates retail gasoline price's cumulative response after a one-unit change in the spot market price for the five periods determined by structural break tests discussed in Section 5.2. The 95% confidence intervals are obtained with bootstrapping methods.



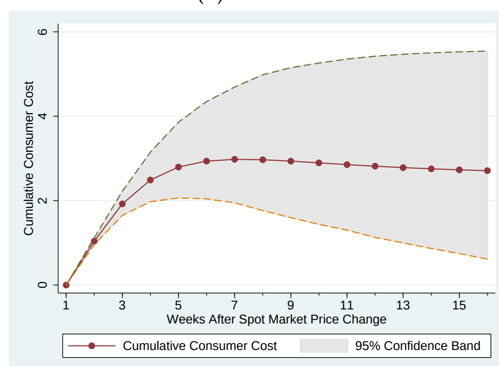
(a) Period 1



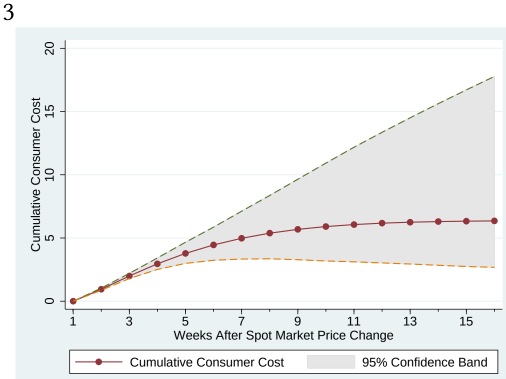
(b) Period 2



(c) Period 3



(d) Period 4



(e) Period 5

Figure 3.5: Consumer Costs for Five Structural Break Periods

This figure illustrates the cumulative consumer costs caused by asymmetric adjustment speeds for the five periods determined by structural break tests discussed in Section 5.2. The 95% confidence intervals are obtained with bootstrapping methods.

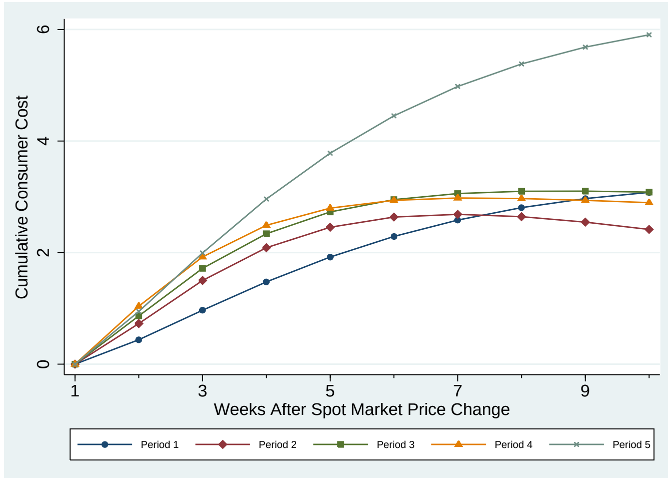


Figure 3.6: Consumer Costs Comparison Across Five Structural Break Periods

This figure compares the cumulative consumer costs across the five periods determined by structural break tests discussed in Section 5.2. Period 5 (October 2008-April 2013) exhibits significantly higher consumer costs than the first four periods.

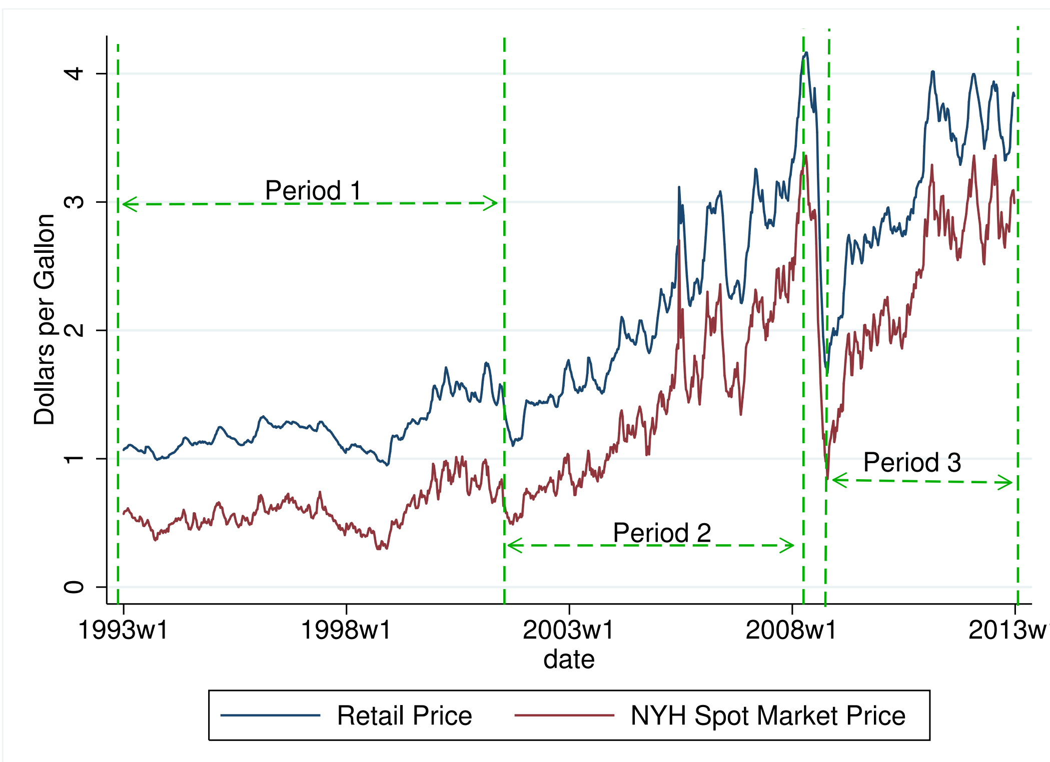
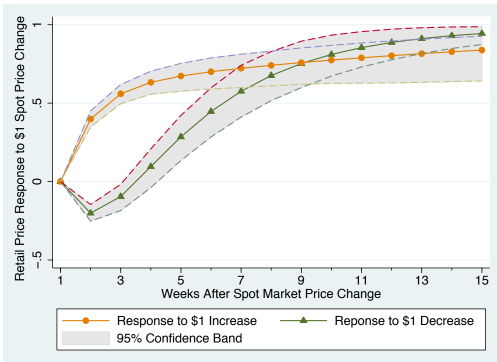
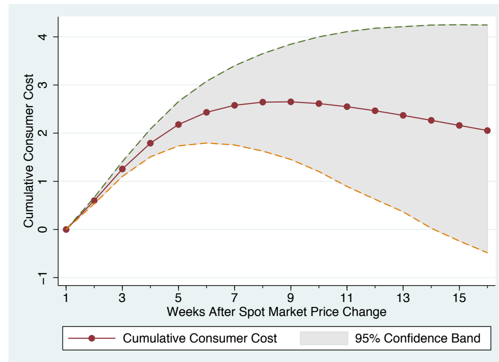


Figure 3.7: Re-partition into Three Periods after Discarding Data from the Great Recession

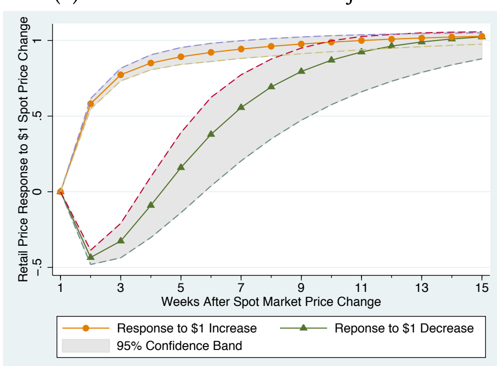
This figure illustrates the second partition of the 20-year sample after dropping data from August 8, 2008 to March 27, 2009, during which gasoline prices decreased rapidly. The three periods shown here are: 1) April 2, 1993-March 16, 2001; 2) March 23, 2001-August 1, 2008; 3) April 3, 2009-March 29, 2013.



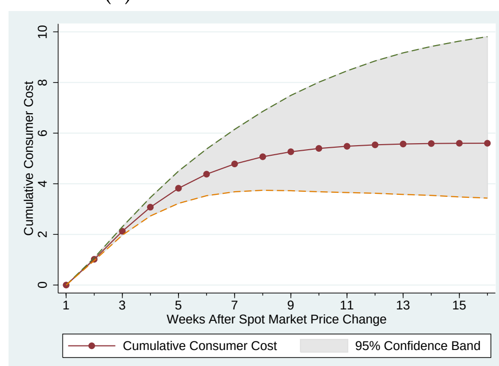
(a) Period 1 Cumulative Adjustments



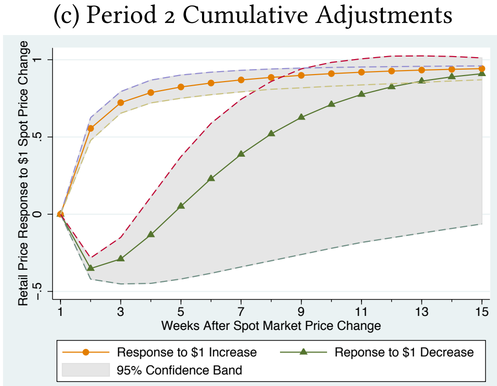
(b) Period 1 Consumer Costs



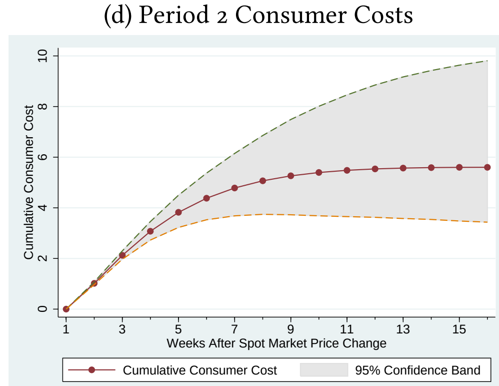
(c) Period 2 Cumulative Adjustments



(d) Period 2 Consumer Costs



(e) Period 3 Cumulative Adjustments



(f) Period 3 Consumer Costs

Figure 3.8: Cumulative Adjustment Functions and Consumer Costs for Three Periods

This figure depicts cumulative adjustment functions and consumer costs for the three periods discussed in Section 5.3. The 95% confidence intervals are obtained with bootstrapping methods.

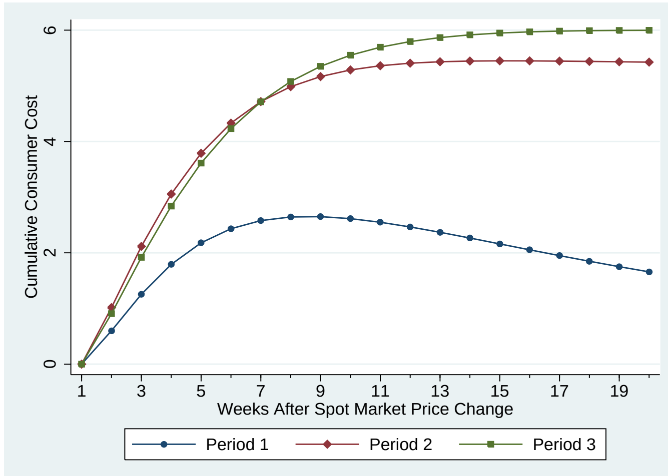
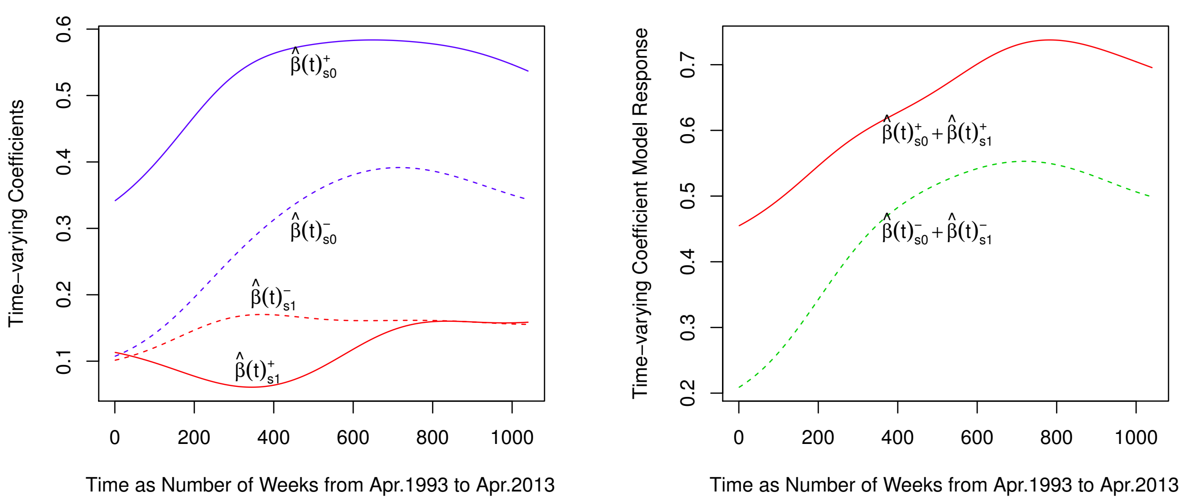


Figure 3.9: Consumer Costs Comparison Across Three Periods

This figure compares the consumer costs across the three periods discussed in Section 5.3. The latter two periods (March 23, 2001-August 1, 2008 and April 3, 2009-March 29, 2013) exhibit higher consumer costs than the first period (April 2, 1993-March 16, 2001).



Panel A: Time-Varying Coefficients

Panel B: Time-Varying Coefficients Combined

Figure 3.10: Time-Varying Coefficients (Combined) for Spot Price Changes

This figure depicts the estimated functions $\beta^{+}_{s0}(t)$, $\beta^{-}_{s0}(t)$, $\beta^{+}_{s1}(t)$, and $\beta^{-}_{s1}(t)$ of the time-varying coefficient model discussed in Section 6. Panel A presents the contemporaneous and previous period responses of the retail gasoline price to spot market price changes separately. Panel B presents the two-period accumulated effects of a positive or negative change in spot price.

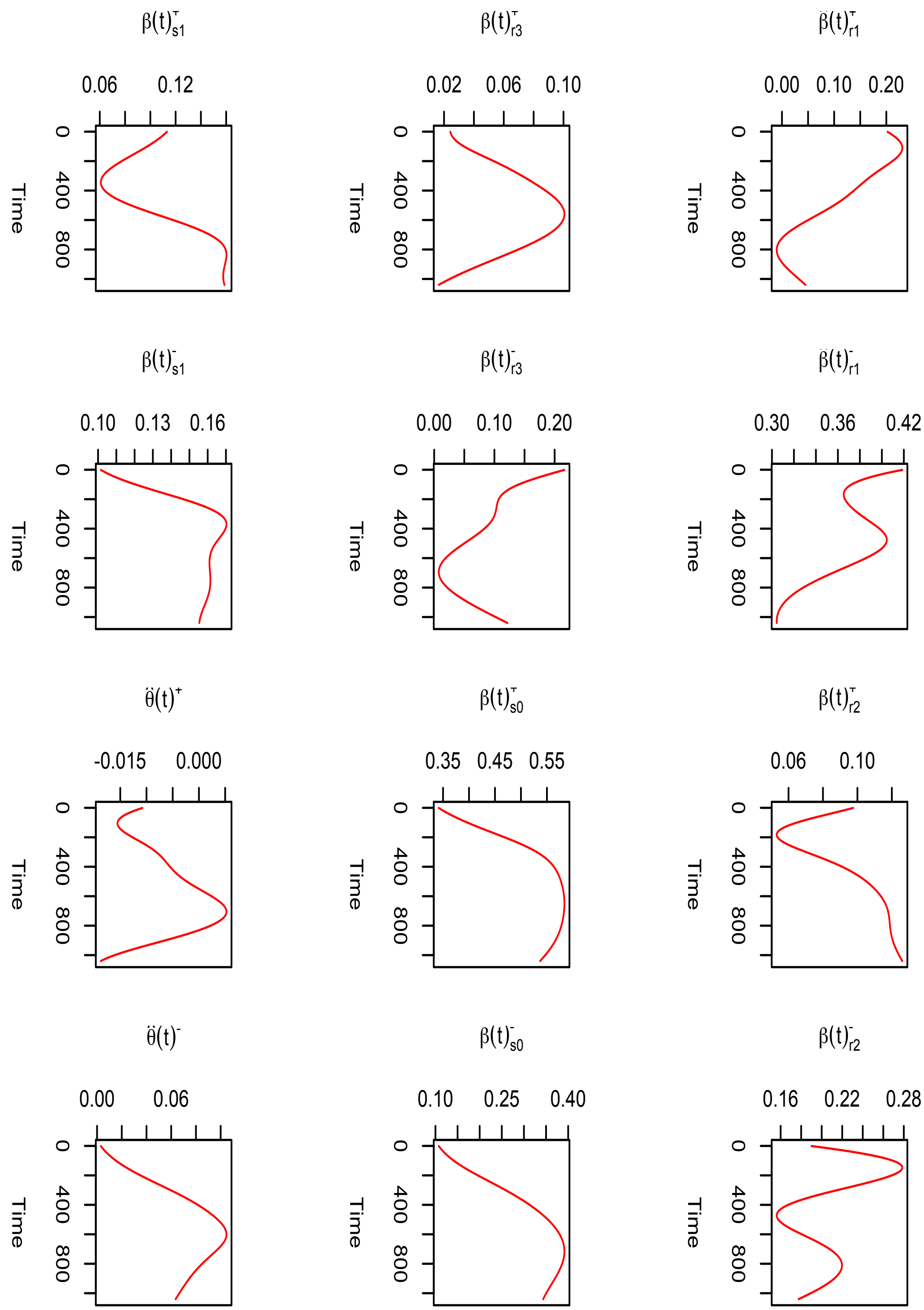


Figure 3.11: Time-Varying Coefficients

This figure exhibits all the estimated coefficient functions of the time-varying coefficient model discussed in Section 6. The number of lags selected by BIC is 3 for the retail prices and 1 for the spot market prices.

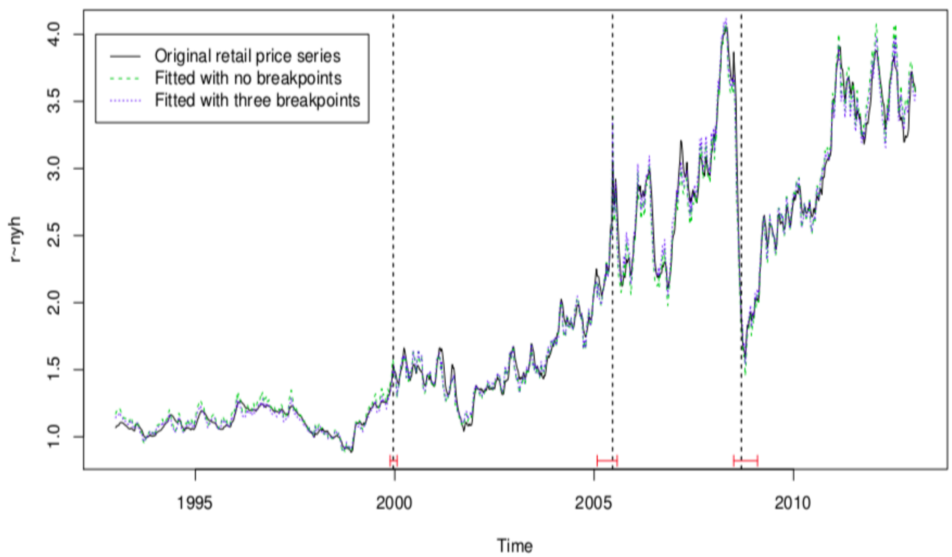


Figure 3.12: Breakpoints for the Long-Run Equilibrium Between Retail and Spot Prices

This figure presents the detected structural breakpoints for the long-run equilibrium between retail and spot market prices. Fitted values with structural breaks are shown in blue dotted lines. Fitted values without structural breaks are shown in the green dashed line. The three detected breakpoints are March 10, 2000, September 2, 2005, and November 21, 2008. The red intervals enclosing each breakpoint are the 95% confidence intervals.

References

[1] S. Agarwal, B. W. Ambrose, S. Chomsisengphet, and C. Liu. The Role of Soft Information in a Dynamic Contract Setting: Evidence from the Home Equity Credit Market. *Journal of Money, Credit and Banking*, 43(4):633–655, 2011.

[2] S. Agarwal and R. Hauswald. Distance and Private Information in Lending. *The Review of Financial Studies*, 23(7):2757–2788, 2010.

[3] S. Agarwal, V. Mikhed, and B. Scholnick. Does the Relative Income of Peers Cause Financial Distress? Evidence from Lottery Winners and Neighboring Bankruptcies. Working Paper 18-16, Federal Reserve Bank of Philadelphia, May 2018.

[4] A. Alexandrov and S. Koulayev. No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information. Working Paper 2017-01, Consumer Financial Protection Bureau Office of Research.

[5] J. Allen, R. Clark, and J.-F. Houde. Search Frictions and Market Power in Negotiated Price Markets. Working Paper 1403, Queen's University, Department of Economics, March 2018.

[6] D. F. Amel, A. B. Kennickell, and K. B. Moore. Banking Market Definition: Evidence from the Survey of Consumer Finances. Finance and Economics Discussion Series, U.S. Federal Reserve Board, 2008.

[7] M. Armstrong, J. Vickers, and J. Zhou. Prominence and Consumer Search. *RAND Journal of Economics*, 40(2):209–233, 2009.

[8] L. J. Bachmeier and J. M. Griffin. New Evidence on Asymmetric Gasoline Price Responses. *Review of Economics and Statistics*, 85(3):772–776, 2003.

[9] R. W. Bacon. Rockets and Feathers: The Asymmetric Speed of Adjustment of U.K. Retail Gasoline Prices to Cost Changes. *Energy Economics*, 13(3):211–218, 1991.

[10] J. Bai. Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics*, 79(4):551–563, 1997.

[11] J. Bai and P. Perron. Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics*, 18(1):1–22, 2003.

[12] N. S. Balke, S. P. A. Brown, and M. Yucel. Crude Oil and Gasoline Prices: an Asymmetric Relationship? *Economic and Financial Policy Review*, Q1:2–11, 1998.

[13] L. Bettendorf, S. A. van der Geest, and M. Varkevisser. Price Asymmetry in the Dutch Retail Gasoline Market. *Energy Economics*, 25(6):669–689, 2003.

[14] S. Borenstein, A. C. Cameron, and R. Gilbert. Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes? *The Quarterly Journal of Economics*, 112(1):305–339, 1997.

[15] S. Borenstein and A. Shepard. Dynamic Pricing in Retail Gasoline Markets. *RAND Journal of Economics*, 27(3):429–451, 1996.

[16] S. Borenstein and A. Shepard. Sticky Prices, Inventories, and Market Power in Wholesale Gasoline Markets. *RAND Journal of Economics*, 33(1):116–139, 2002.

[17] J. J. Brown and P. H. Reingen. Social Ties and Word-of-Mouth Referral Behavior. *Journal of Consumer Research*, 14(3):350–362, 1987.

[18] Z. Brown. An Empirical Model of Price Transparency and Markups in Health Care. Manuscript, 2018.

[19] K. Burdett and K. L. Judd. Equilibrium Price Dispersion. *Econometrica*, 51(4):955–969, 1983.

[20] P. S. Calem and L. I. Nakamura. Branch Banking and the Geography of Bank Pricing. *The Review of Economics and Statistics*, 80(4):600–610, 1998.

[21] A. Cavallo, B. Neiman, and R. Rigobon. Currency Unions, Product Introductions, and the Real Exchange Rate. *The Quarterly Journal of Economics*, 129(2):529–595, 2014.

[22] A. Cavallo, B. Neiman, and R. Rigobon. The Price Impact of Joining a Currency Union: Evidence from Latvia. *IMF Economic Review*, 63(2):281–297, September 2015.

[23] S. DellaVigna and M. Gentzkow. Uniform Pricing in U.S. Retail Chains. Working Paper 23996, National Bureau of Economic Research, November 2017.

[24] G. Deltas. Retail Gasoline Price Dynamics and Local Market Power. *The Journal of Industrial Economics*, 56(3):613–628, 2008.

[25] G. Deltas and K. Desmet. Welfare Costs of Price Discrimination: Evidence from Uniform Pricing in Currency Unions. Working paper, 2018.

[26] R. Engle and C. Granger. Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2):251–276, 1987.

[27] M. Galeotti, A. Lanza, and M. Manera. Rockets and Feathers Revisited: an International Comparison on European Gasoline Markets. *Energy Economics*, 25(2):175–190, 2003.

[28] U. G. Gurun, G. Matvos, and A. Seru. Advertising Expensive Mortgages. *The Journal of Finance*, 71(5):2371–2416, 2016.

[29] J. D. Hamilton and M. C. Davis. Why Are Prices Sticky? The Dynamics of Wholesale Gasoline Prices. *Journal of Money, Credit and Banking*, 36(1):17–37, 2004.

[30] G. Hitsch, A. Hortaçsu, and X. Lin. Prices and Promotions in U.S. Retail Markets: Evidence from Big Data. Working Paper No.17-18, Chicago Booth Research Paper, November 2017.

[31] A. Hortacsu and C. Syverson. Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds. *The Quarterly Journal of Economics*, 119(2):403–456, 2004.

[32] D. O. S. II. Oligopolistic Pricing with Sequential Consumer Search. *The American Economic Review*, 79(4):700–712, 1989.

[33] W. Jiang, A. A. Nelson, and E. Vytlacil. Liar's Loan? Effects of Origination Channel and Information Falsification on Mortgage Delinquency. *The Review of Economics and Statistics*, 96(1):1–18, March 2014.

[34] R. N. Johnson. Search Costs, Lags and Prices at the Pump. *Review of Industrial Organization*, 20(1):33–50, 2002.

[35] B. Keys, T. Mukherjee, A. Seru, and V. Vig. Financial Regulation and Securitization: Evidence from Subprime Loans. *Journal of Monetary Economics*, 56:700–720, July 2009.

[36] B. Keys, T. Mukherjee, A. Seru, and V. Vig. Did Securitization Lead to Lax Screening? Evidence from Subprime Loans. *The Quarterly Journal of Economics*, 125(1):307–362, 2010.

[37] B. Keys, A. Seru, and V. Vig. Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets. *The Review of Financial Studies*, 25(7):2071–2108, 2012.

[38] Y. S. Kim, S. M. Laufer, K. Pence, R. Stanton, and N. Wallace. Liquidity Crises in the Mortgage Market. Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System (U.S.), March 2018.

[39] J. Lacko and J. Pappalardo. Improving Consumer Mortgage Disclosures. Bureau of Economics Staff Report, U.S. Federal Trade Commission, 2007.

[40] W. W. Lang and L. I. Nakamura. A Model of Redlining. *Journal of Urban Economics*, 33(2):223–234, 1993.

[41] M. S. Lewis. Asymmetric Price Adjustment and Consumer Search: An Examination of the Retail Gasoline Market. *Journal of Economics & Management Strategy*, 20(2):409–449, 2011.

[42] R. Manning and P. Morgan. Search and Consumer Theory. *The Review of Economic Studies*, 49(2):203–216, 1982.

[43] R. Manning and P. Morgan. Optimal Search. *Econometrica*, 53(2):923–944, 1985.

[44] R. S. McMillan. Different Flavor, Same Price: The Puzzle of Uniform Pricing for Differentiated Products. Manuscript, January 2007.

[45] J. Mondragon. Household Credit and Employment in the Great Recession. Nielsen Dataset Paper Series 1-025, Kilts Center for Marketing at Chicago Booth, March 2018.

[46] S. Peltzman. Prices Rise Faster than They Fall. *Journal of Political Economy*, 108(3):466–502, 2000.

[47] J. Perdiguero-Garciá. Symmetric or Asymmetric Oil Prices? A Meta-analysis Approach. *Energy Policy*, 57:389–397, 2013.

[48] M. A. Petersen and R. Rajan. Does Distance Still Matter? The Information Revolution in Small Business Lending. *Journal of Finance*, 57(6):2533–2570, 2002.

[49] A. F. Presbitero and A. Zazzaro. Competition and Relationship Lending: Friends or Foes? *Journal of Financial Intermediation*, 20(3):387–413, 2011.

[50] S. Radchenko. Oil Price Volatility and the Asymmetric Response of Gasoline Prices to Oil Price Increases and Decreases. *Energy Economics*, 27(5):708–730, 2005.

[51] L. Relihan. Branches in Local Mortgage Markets. Working paper, 2017.

[52] R. J. Rosen. Competition in Mortgage Markets: The Effect of Lender Type on Loan Characteristics. *Economic Perspectives*, 35(1), 2011.

[53] S. Salop and J. Stiglitz. Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion. *The Review of Economic Studies*, 44(3):493–510, 1977.

[54] T. Salz. Intermediation and Competition in Search Markets: An Empirical Case Study. Working Paper, 2015.

[55] D. Scharfstein and A. Sundaram. Market Power in Mortgage Lending and the Transmission of Monetary Policy. Working Paper, 2014.

[56] B. Shiller and J. Waldfogel. Music for a Song: An Empirical Look at Uniform Pricing and Its Alternatives. *Journal of Industrial Economics*, 59(4):630–660, 2011.

[57] A. T. Sorensen. Equilibrium Price Dispersion in Retail Markets for Prescription Drugs. *Journal of Political Economy*, 108(4):833–850, 2000.

[58] G. J. Stigler. The Economics of Information. *Journal of Political Economy*, 69(3):213–225, 1961.

[59] H. R. Varian. A Model of Sales. *The American Economic Review*, 70(4):651–659, 1980.

[60] W. H. Whyte. The Web of Word of Mouth. *Fortune*, 50:140–143, November 1954.

[61] L. L. Wilde and A. Schwartz. Equilibrium Comparison Shopping. *The Review of Economic Studies*, 46(3):543–553, 1979.

[62] S. E. Woodward and R. E. Hall. Diagnosing Consumer Confusion and Sub-optimal Shopping Effort: Theory and Mortgage-Market Evidence. *American Economic Review*, 102(7):3249–3276, 2012.

[63] A. Zeileis, C. Kleiber, W. Kramer, and K. Hornik. Testing and Dating of Structural Changes in Practice. *Computational Statistics and Data Analysis*, 44(1):109–123, 2003.